Mutations in the Charged Domain of CBX2 Disrupt PRC1 Function in Vivo

Mutations in the charged domain of CBX2 disrupt PRC1 function *in vivo*

A dissertation presented

by

Mei Sheng Lau

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Genetics

Harvard University

Cambridge, Massachusetts

April 2016

Dissertation Advisor: Dr. Robert E. Kingston Mei Sheng Lau

Mutations in the charged domain of CBX2 disrupt PRC1 function *in vivo*

ABSTRACT

Epigenetics inheritance is a phenomenon where a cell state is inherited across cellular divisions. It is important for maintaining differentiated cells types, and thus is fundamental to the development of multicellular organisms. Failure in the system can lead to deleterious effects that range from embryonic lethality to the development of diseases such as cancer.

The Polycomb group (PcG) proteins are important regulators of epigenetic inheritance: they maintain the repression of the *Hox* genes, which specify the body axes of an organism, throughout development and into adulthood. The loss of function of PcG genes therefore leads to homeotic transformations. A mechanistic understanding of PcG function will yield insight into the mechanisms involved in epigenetics inheritance.

The Polycomb Repressive Complex 1 (PRC1) is able to inhibit chromatin remodeling and compact polynucleosomes *in vitro*; these activities could represent how stable gene repression is achieved *in vivo.* However, it has been difficult to determine the biological relevance of these activities because the nature of nucleosomal-level compaction in the nucleus is unknown and there is no direct assay for it.

To circumvent this problem, we adopted an approach where we introduced into mouse embryonic stem cells (mESCs) and mice the mutations that disrupt the *in vitro* activities. The mutations are specific lysine-to-alanine and arginine-to-alanine substitutions in a positively charged domain of the CBX2 subunit of PRC1. These are known to disrupt

PRC1 polynucleosmal compaction activities from previous studies. We compared the mutant mESCs and mice to wild type counterparts for PcG-related phenotypes.

We observed that the mutations prevented the repression of CBX2 (also PRC1) target genes in mESCs. They did not prevent CBX2 from forming PRC1 or binding to chromatin. This suggests that the *in vitro* activities are specifically involved in inhibiting transcription.

Moreover, we observed that the *Cbx2* mutant mice exhibit the classic PcG phenotype, which is homeotic transformation in the anterior-posterior axis. The posterior transformations are similar to those observed in the *Cbx2*$^{-/-}$ mice, indicating that our mutations mimic loss-of-function of *Cbx2.* Notably, the *Cbx2*$^{-/-}$ mice had other phenotypes in addition to the PcG phenotype, which we did not observe in our mutants. This indicates that the mutations in the compaction domain of CBX2 specifically disrupted a PcG function during development.

These experimental outcomes are consistent with the hypothesis that the *in vitro* compaction activity of PRC1 is important for how heritable gene repression can be achieved *in vivo.*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to dedicate this dissertation to my advisor Dr. Robert Kingston. This dissertation reports the first genetics experiment that was performed in his laboratory, despite Bob being the Vice-Chair of the Department of Genetics all these years. Thank you, Bob, for your example, support and encouragement to reach beyond personal limits.

I have benefited from the coaching of several lab members: Aimeé Deaton, my former bay mate who read and commented on almost everything I had written about science in the past four years; Sharmistha Kundu, who gave me my first formal lesson on giving scientific presentations (for PQE), and had since taught me much more about doing science; Patrick Schorderet, who shared with me his extensive knowledge and enthusiasm about developmental biology, without which the work in Chapter 3 would have taken a more arduous path to complete.

I want to acknowledge the tremendous help from the following people, without whom my experience with mouse work could have been traumatizing: William Press, who taught me how to handle the furry creatures from the start; Sara Miller and Jason West, who worked hard to set up the mouse protocol; Sharon Marr, whose help with genotyping spared me from doing PCRs all day long.

I have the good fortune of meeting many wonderful people who stood with me through the good and bad times of graduate school: Papa Jerry and Mama Judy Ball, Pastors

Chris and Ishtar Gleason, Jofeanne Gaco and Raquel Mota-Hays. I believe I will only know the full impact of your prayers and intercessions for me in Heaven.

I must also mention my parents Weng Ann Lau and Lung Meng Tang, and my brothers John and Henry Lau, who have not ceased to show their love and care for me in spite of the 12 to 13 hour time difference and immense physical distance between us.

Finally, I want to give all glory to my Lord and Savior Jesus Christ. Again and again, He has proven to me that He is El Shaddai, God Almighty. There is nothing too hard and too wonderful for Him.

"Blessed is the man (woman)…
who delights in the law of the LORD.
He (She) is like a tree
planted by streams of water
that yields its fruit in its season,
and its leaf does not wither.
In all that he (she) does, he (she) prospers. "

-Psalm 1:1-3

CHAPTER 1

INTRODUCTION

Epigenetic inheritance

The formation of a multicellular organism during embryogenesis is a fascinating process. A single fertilized egg undergoes repeated rounds of cellular division, and the daughter cells become progressively specified, committed and eventually fully differentiated into the diverse cell types that make the organism. Cellular division and differentiation occur concurrently, which necessitates that an acquired cell fate to be passed on to the daughter cells.

The signals that induce cell fate specification during embryogenesis are often transient, yet the cell fates are maintained throughout the developmental process and life of the organism. This implies that there are cellular memory mechanisms in place that allow acquired cell states to be maintained in the absence of external signals and to be propagated through cellular divisions. This cellular memory is a form of epigenetic inheritance.

Many cell fates are determined through the activation and repression of specific gene networks in the cell. Epigenetic inheritance thus involves maintaining the gene expression patterns that have been established. The Polycomb group (PcG) genes are a class of regulators that are important in this process.

The stability of a cell fate is clearly important. Mutations in the components that regulate this process, including the in the *PcG* genes, lead to developmental defects, some of

which are embryonic lethal. Mutations in these components in adult tissues are associated with a wide range of cancers[1]. Therefore, understanding the mechanism of epigenetic inheritance is a central problem in biology. This dissertation focuses on the roles of the *PcG* genes as epigenetic regulators, specifically on their function in achieving heritable gene repression.

PcG genes are epigenetic regulators that are widely conserved

The PcG genes are epigenetic regulators because they are required for maintaining *Hox* gene repression in specific body segments throughout embryogenesis and in adulthood[2,3]. The *Hox* genes encode a class of transcription factors whose collective activity is an important determinant of the anterior-posterior axis in almost all metazoans[4,5]. The *Hox* genes are arranged in clusters in the genome, and the arrangement is closely linked to their function. The regulation of the *Hox* genes exhibit the curious pattern of spatial, temporal and functional collinearity: their expression is sequentially activated from the 3' to 5' end of the cluster during embryogenesis (spatial and temporal collinearity), which in turn specify segment identities along the body axis (functional collinearity) (Figure 1.1). In flies, the *Hox* gene expression pattern along the body axis is initiated by early developmental factors[6,7]. These factors are transiently expressed, but the expression pattern persists throughout development into adulthood. Perturbations to the individual *Hox* gene expression boundaries affect the body plan of the animal, which can manifest in the embryonic, larval or adult stages as homeotic transformations in the anterior-posterior axis.

The *PcG* genes were initially identified in mutant flies with ectopic sex combs on their second and third legs[3,8]. As sex combs are usually exclusively on the first legs, the phenotypes suggested there was homeotic transformation in the anterior-posterior axis of the animals. Some PcG mutants have segmentation defects in their embryos and larvae, further corroborating that the PcG genes are involved in proper patterning of the anterior-

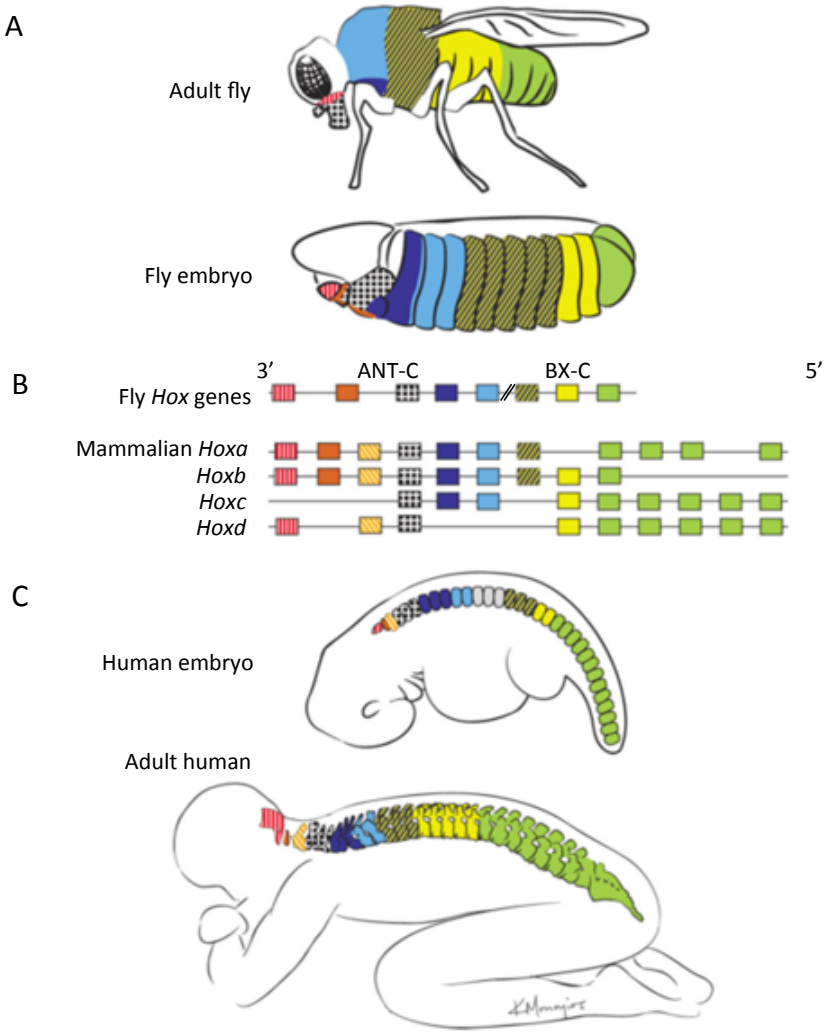


Figure 1.1 Spatial and functional collinearity of the *Hox* genes.
B. The genomic organization of the *Hox* clusters in the fly and mammalian (human) genomes. There are four paralogous *Hox* clusters in mammals. The genes that are orthologous between clusters and species have the same color. The genes are sequentially activated according to their positions in the cluster, starting from the 3'end. **A. and C.** Schematic illustrations of the anterior-posterior axes in the embryos and adults of flies and humans, respectively. The color codes correspond to the expression boundaries of the *Hox* genes indicated in panel B. The *Hox* gene expression patterns confer positional identities along the axes. The patterns are established in the early developmental stages and maintained through embryogenesis to adulthood, as depicted. The figure is adapted from www.yourinnerfish.com.

posterior axis[3,4]. Through examining PcG mutant embryos at early and later stages of development, it was found that the *Hox* gene boundaries were appropriately set up but spread into more anterior segments in the later embryo[2,9]. This indicated that PcG genes are required to prevent ectopic *Hox* gene expression to maintain the pattern that was established during early development. Hence, PcG genes are epigenetic regulators that maintain gene repression. The PcG consists of many members; of particular relevance to this dissertation are *Polycomb (Pc), Posterior sex comb (Psc), Polyhomeotic (Ph) and Sex comb extra (Sce).*

Homologs to the fly PcG genes have been identified in animals and plants by virtue of sequence identities in the genes, the presence of conserved protein domains and/or functional conservation. In mammals, the earliest PcG genes known were the homologs to *Pc, Cbx2* (also known as *M33*), *the* homologs to *Psc, Bmi-1* and its paralog *Mel-18*, and the homolog to Ph, Phc1 (also known as Mph1)[10-13]. Knockout mouse mutants of *Cbx2, Bmi-1, Mel-18* and *Phc1* all lead to posterior transformation in the axial skeleton, suggesting that PcG regulation of the *Hox* genes during development is conserved between flies and mice. As in the case for the fly mutants, the posterior transformations in the mice were accompanied by anterior shifts the expression boundaries of several *Hox* genes. In addition, the expression of mouse *Cbx2* can partially rescue the *Pc* mutant phenotype in transgenic flies[14]. These observations imply that the molecular mechanism for epigenetic repression by PcG is widely conserved.

The mammalian PcG system, however, is much more complex than its fly counterpart. There are multiple homologs for almost every fly PcG gene[15,16]. The paralogs have homologous domains and variable domains, and they tend to be differentially regulated during embryogenesis[17,18]. This suggests that they have both common and non-overlapping functions. Indeed, knockout mutants of different paralogs have shared and

distinct phenotypes. For example, *Bmi*$^{-/-}$ and *Mel-18*$^{-/-}$ mice both have posterior transformations that corresponded to anterior shifts in *Hox* expression, but only *Bmi*$^{-/-}$ exhibited neurological abnormalities while *Mel-18*$^{-/-}$ had intestinal defects [11,12]. Of note, the distinct phenotypes many or may not be related to the PcG function. In some instances, knocking-out certain mammalian PcG homologs do not lead to anterior-posterior transformations, for example in *Cbx7*$^{-/-}$ and *Cbx8*$^{-/-}$, even though the molecular evidence (e.g. protein-protein interactions and chromatin binding sites) strongly suggests that they are members of the PcG system[19-22]. The absence of phenotype could be due to functional redundancy from another paralog, or that the molecular features commonly associated with PcG may not always be related to PcG function. Therefore, studying the mammalian PcG system has its additional challenges.

In view of the integral role that PcG genes play in cell fate maintenance across species, an understanding of the mechanistic details of their activities will prove insightful to how cell fates are stabilized and propagated through cellular divisions. In the following sections, we review the molecular characterization of the PcG gene products.

Polycomb group protein complexes and chromatin

The PcG genes encode proteins that form several multi-subunit complexes. The various types of PcG complexes are illustrated in Figure 1.2. Among these, two complexes, namely Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (PRC2), are the best characterized. Their compositions are conserved between flies and mammals. As alluded to in the previous section, there can be variation in the composition of the mammalian PRC1 and PRC2 as there can be different combinations of paralogs that form them.

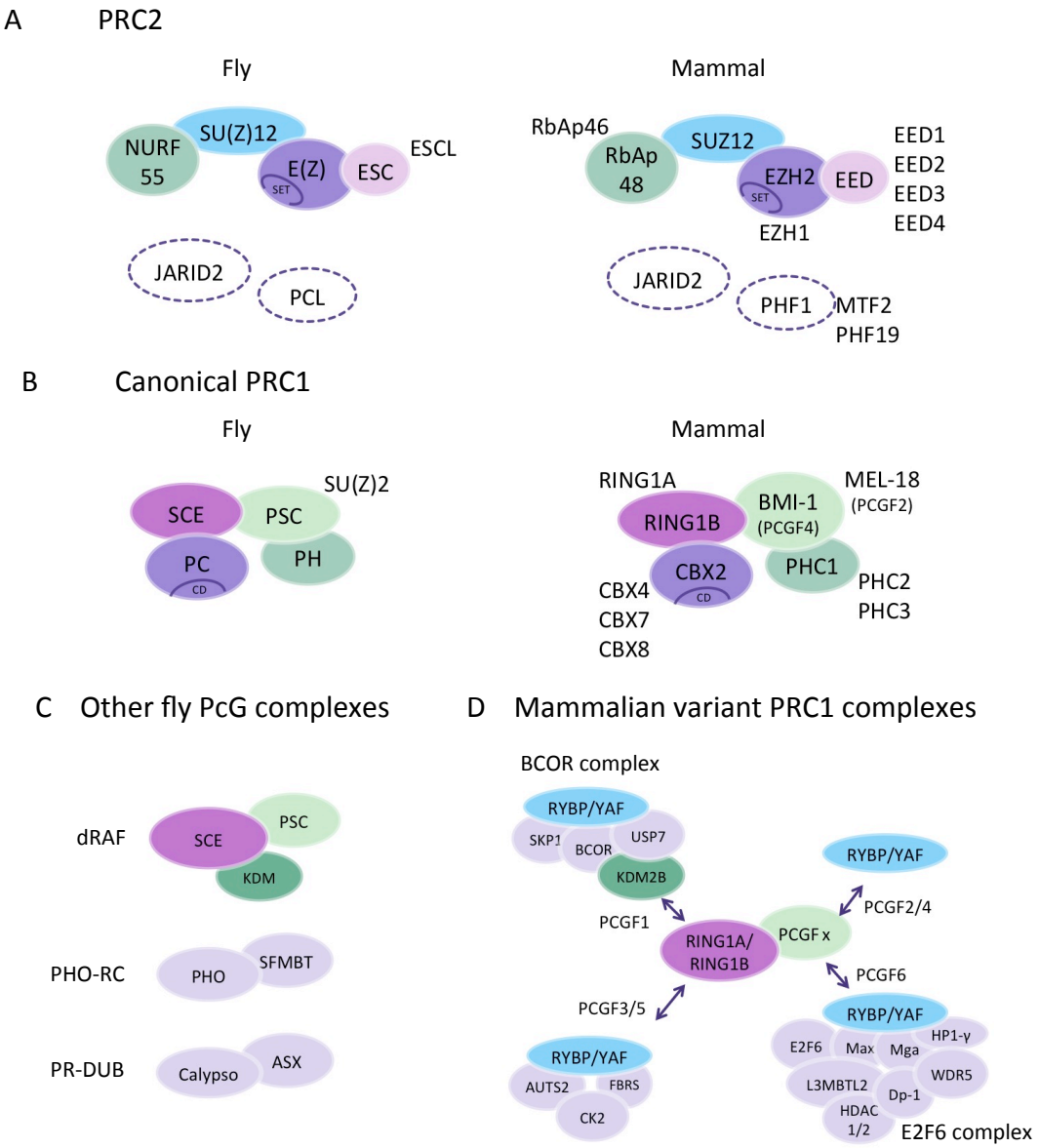


Figure 1.2 Fly and mammalian PcG complexes and their subunits.
The filled ovals depict the core subunits. The open ovals with dashed borders depict ancillary subunits. The text next to the ovals indicate the paralogs of the subunit that can be substituted into the complex. **A.** PRC2 from flies and mammals. The SET domain of E(Z) and EZH2 are indicated. **B.** Canonical PRC1 from flies and mammals and their four core subunits. The chromodomain (CD) of PC and its homolog (CBX2) are shown. **C**. Other known fly PcG complexes. **D.** The family of PRC1 variant complexes in mammals. The complexes all have Ring1A/B and a PCGF paralog as core subunits. These complexes contain RYBP or YAF proteins, in place of a CBX paralog which is found exclusively in the canonical PRC1. The figures are adapted from Simon and Kingston (2013) and Aranda *et. al* (2015).

Several early observations in flies had suggested that the PcG proteins exert their gene silencing effects through modulating chromatin. PcG proteins localized at many discrete loci on polytene chromatin from fly salivary glands. These sites included their known target genes, the *Hox* genes in the *Antennapedia* (ANT-C) and *Bithorax* Complexes (BX-C). The overexpression and loss-of-function of PcG genes lead to enhanced polytene chromosome condensation and decondensation, respectively[23,24].

The fly PC protein is a core component of PRC1 and has an N-terminal domain (later named 'chromodomain') that resembled the chromatin-binding domain of Heterochromatin Protein 1 (HP1)[25]. HP1 is important for position effect variegation (PEV). It associates with heterochromatin, which appears as dense structures under the light microscope. As PEV also involves propagating a repressed transcriptional state across cell divisions, it was thought that PC could maintain *Hox* gene repression through similarly creating higher order chromatin structures. While the chromodomain of HP1 recognizes and binds to the trimethylated lysine 9 of histone H3 (H3K9me3), the chromodomain of PC recognizes and binds to the trimethylated lysine 27 of histone H3 (H3K27me3). H3K27me3 is a mark that is deposited by PRC2. The localization sites PRC1, PRC2 and H3K27me3 on chromatin highly overlap genome-wide[26-28]. This led to a model where PRC2 first tri-methylates histone H3, which is recognized by the chromodomain of PC and thus recruits PRC1 to the target genes. PRC1 then inhibits transcription by its activities on the chromatin, as discussed in the next section.

The PcG phenotype is suppressed by mutations in the *trithorax* group (trxG) genes[29]. The *trxG* group genes encode the subunits of SWI/SNF, an ATP-dependent chromatin remodeler that is required for sustained *Hox* gene expression. The opposing functions of the PcG and trxG on *Hox* gene expression suggested that they have direct antagonistic effects on

chromatin for creating prohibitive and permissive chromatin states, respectively, for transcription regulation.

Complementary approaches to understanding PcG-mediated gene repression

In view of the above observations, biochemical assays that aim to determine how PcG proteins might influence chromatin were developed. These assays were made possible by technical breakthroughs in two areas: firstly, the expression and purification of PRC1 and PRC2 subunits, and the reconstitution of the minimal complexes with recombinant proteins; secondly, the assembly of nucleosomes on DNA with specific nucleotide sequences that serve as *in vitro* chromatin template. The robustness of the assays had enabled investigators to isolate not only the subunits, but also identify the domain or the specific amino acid residues within the subunits, that are required for an activity. These studies provided insight into the mechanistic details of how PRC1 and PRC2 may influence their target genes in the nucleus.

Nonetheless, the caveat to the biochemical assays is that they often involve non-physiological conditions. For instance, an assay may involve non-physiological salt conditions, which may generate protein behaviors that are non-existent *in vivo*. Hence, their relevance must at least be tested in a cellular system with molecular assays.

In a cellular system, transcriptional silencing of PcG protein-bound genes is used as the readout for PcG function. Chromatin immunoprecipitation (ChIP) experiments followed by quantitative PCR (qPCR) or high throughput sequencing analyses are routinely used to identify the genes that are bound by a PcG protein of interest. Investigators perturb the system by depleting, overexpressing or mutating the PcG protein and assay for gene expression changes. As PcG proteins exist as multi-subunit complexes *in vivo*, the depletion or overexpression of a PcG protein often affect other components of the complex (for

example, see Ref. 30) Hence, it is not possible to establish direct relationship between an activity and gene repression through these experiments. With the knowledge gained from biochemical and structural studies, once can selectively mutate the key residues required for the activity without disrupting the complex, thereby isolating the effect of the activity on PcG function.

The test for the relevance of a biochemical activity in cells may or may not be straightforward. In the case of PRC2, its histone methyltransferase (HMTase) activity, which catalyzes the tri-methylation of histone H3 lysine 27 (H3K27me3), was readily assayed for in cells using antibodies and found to be associated with PcG repression[31]. Furthermore, the SET domain of E(z), which catalyzes the modification, is conserved and well-described from previous studies. Through selective mutations at the critical residues that were identified by homology to other HMTase proteins, catalytic dead PRC2 mutants were generated and found to have PcG phenotypes. Thus, for PRC2, the activity-to-function relationship was rapidly established. Non-enzymatic activities, however, are more challenging to assay for in cells. Thus, it can be complicated to determine if they take place *in vivo* and /or have biological functions. This is especially so if they involve protein regions with no apparent homology to other known proteins. The polynucleosome compaction activity of PRC1, which depends on an intrinsically unfolded region of CBX2, is one such example, which we will discuss in greater detail in the next section.

Overall, we have gained much knowledge about the PcG system since the identification of the first PcG gene, *Pc*, in the year 1947[8]. This was through biochemical, molecular and genetic studies that provided complementary information. In the following section we describe three examples of PRC1 activities and our current understanding of how they may contribute to the role of PRC1 as an epigenetic gene repressor. Our focus is on the mammalian PRC1, though the activities are conserved in the fly PRC1. We will also

highlight a key gap in the prevailing knowledge, which we aim to address with the research work presented in this dissertation.

Mechanisms of gene repression by PRC1

PRC1 consists of four core components, which are conserved between fly and mammals: an SCE homolog, a PSC homolog, a PC homolog and a PH homolog (Figure 1.2)[32,33]. The subunits have distinct biochemical activities, which imply that the complex may repress its target genes through multiple mechanisms. It was observed that genes that are bound by all four PRC1 core subunits in flies fall into two broad categories: Class I genes, which include the *Hox* genes, require all four components for repression, while Class II genes depend only on PSC and PH[34]. This indicates that there may be differential and synergistic regulation at different target genes. Here, we review the following three activities of PRC1:

1. Histone H2A mono-ubiquitination by the SCE/PSC homologs
2. Chromatin organization by the PH homologs
3. Nucleosomal-level compaction by fly PSC and mammalian CBX2

1. Histone H2A mono-ubiquitination by the SCE/PSC homologs

Wang *et. al* (2004) discovered that the RING1B (or RING1A)/BMI-1 complex purified form HeLa cells is a E3 ubiquitin ligase enzyme that mono-ubiquitylates histone H2AK119[35]. This catalytic activity is conserved in the homologous fly SCE/PSC complex. The mono-ubiquitynated histone H2A (H2AUb) is localized at the promoters of PRC1 target genes in HeLa and fly S2 cells. The depletion of Ring1B or SCE leads to loss of H2AUb and ectopic expression of the target genes. In addition, H2AUb is enriched on the inactive X-chromosome[36-38].These observations suggested that H2AUb is involved in stable gene

repression. Notably, the RING1B/BMI-1 enzymatic module is not exclusive to the canonical PRC1, but is also present in several PRC1-like complexes such as BCOR and E2F6 in mammals, and dRAF in flies (Figure 1.2)[16,39].

The role of H2AUb in PRC1-mediated repression is unclear. H2AUb was thought to be required for restraining poised RNA polymerases, at least on bivalent genes in mouse embryonic stem cells (mESCs)[40]. However, there is increasing evidence showing that it is mainly involved in PcG protein recruitment in mammals. Through using a *de novo* targeting assay in mESCs, two groups observed that H2AUb is required for PRC2, and thus H3K27me3, localization on target genes[41,42]. The H3K27me3 in turn recruits canonical PRC1 via interactions with its chromodomain-containing subunit. As canonical PRC1 also has the H2A ubiquitylating activity, it was proposed that this cascade of events forms a feedback loop that reinforces PRC1 and PRC2 binding to create a stable repressive domain.

The protein complex that deposits the initiating H2AUb modifications, however, is a variant PRC1 that contains Rybp and/or Kdm2b[41,43]. Unlike the canonical PRC1, the variant complex does not contain the PC and PH homolog subunits. It may also contain an alternative PSC homolog, PCGF1 (Figure 1.2). This is consistent with a previous finding where the depletion of a canonical PRC1 component Bmi-1 does not affect H3K27me3[44].

The variant PRC1 does not appear to be directly involved in the classically defined functions of PRC1. In mESCs, genes that bound by the variant complex are moderately expressed; in contrast, those bound by the canonical PRC1 are repressed[45]. In addition, the target genes of the variant complex are mainly metabolic and cell-cycle regulators, whereas those of the canonical complex are lineage-specific developmental regulators. Thus, it was unclear whether the H2A ubiquitylating activity of the RING1B/BMI-1 module is a critical function of canonical PRC1.

Two groups addressed this question by generating catalytic dead mutants of Ring1B and SCE in mammals and flies, respectively[30,46]. The targeted impairment of the E3 ligase activity was achieved through an I53A substitution, which was shown by X-ray crystallography to lie in the interface that binds the E2 ubiquitin-conjugating enzyme[47]. In both mice and flies, the catalytic function of Ring1B is dispensable for the repression of *Hox* genes and for proper anterior-posterior patterning. Therefore, H2AUb is not required for the classic PcG function. The mutation, however, is embryonic lethal, which indicates H2AUb has other regulatory roles that remain to be explored.

The study of histone H2A mono-ubiquitination by Ring1B is thus an example of how PcG and non-PcG functions of a protein can be dissected through combining biochemical, molecular and genetic approaches

2. Chromatin organization by the PH homologs

The fly PH and all its mammalian homologs have a sterile alpha motif (SAM) domain that has self-polymerizing activity. A crystal structure of the PH-SAM domain showed that it forms a helical polymer, which resembles that formed by another known transcriptional repressor TEL[48]. The similarities prompted a model where gene repression can be achieved through protein-protein interactions that structurally organize the chromatin in space. It was later observed that mouse PRC1 is able to recruit nucleosomal arrays in *trans* and inhibit chromatin remodeling on the recruited array. This activity is dependent on the Ph subunit[49]. This generated further interest in the model where higher order chromatin organization is involved in gene repression.

The hypothesis was tested in cells by assaying chromatin for long-range interactions and organizational features that are dependent on PRC1. Through information collected from conventional immunofluorescence imaging, super-resolution STORM imaging,

chromosome conformation capture assays and chromatin polymer modeling, it was found that chromatin with PRC1 binding do form subnuclear clusters with a unique compact packaging configuration[50,51]. The depletion of Ph and mutation the SAM domain that prevent polymerization both disrupt the clusters and abrogate chromatin interactions, which often coincide with ectopic expression of PRC1 target genes[51,52]. It is, however, unclear how transcription is inhibited within these higher order structures. They may be involved in excluding the transcriptional machinery from the chromatin, isolating the repressed regions from neighboring active genes to prevent 'bystander' transcriptional activation, or stabilizing local PRC1 binding by mass action to enable robust silencing. Nonetheless, the molecular assays in cells show that Ph polymerization is important *in vivo* and is required for gene repression by PRC1.

Finally, a Phc2-SAM mutant mouse model was generated and found to have the classic PcG phenotype of anterior-to-posterior transformation[50]. The mouse embryonic fibroblasts derived from the mutant mice have *Hox* gene mis-expression.

Therefore, the biochemical, molecular and genetic experiments have altogether showed that chromatin organization through protein-protein interactions is an important feature in PRC1-mediated gene repression and PRC1 function during development.

3. Nucleosomal-level compaction by fly PSC and mammalian CBX2

In the year 1999, PRC1 was successfully purified from fly embryos. It was able to prevent ATP-dependent chromatin remodeling by the human SWI/SNF complex on a plasmid template containing assembled nucleosomes[32]. This provided the molecular evidence for the genetic interactions between the *PcG* and *trxG* genes, which had suggested that PcG and trxG complexes have direct antagonistic effects on the *Hox* genes. It was later found that the complex is able to inhibit transcription from a chromatin template *in vitro*,

again providing molecular evidence for the function of PRC1 as a transcriptional regulator[53]. The two biochemical activities correlated well with a third *in vitro* observation, that is the ability of PRC1 to compact polynucleosome arrays: following incubation with PRC1, the conformation of a polynucleosome array, when viewed by electron microscopy, becomes converted from an extended beads-on-a-string structure into a series of dense globules consisting of multiple adjacent nucleosomes[54]. Based on these activities, it was hypothesized that the fly PRC1 could stably repress its target genes through creating a compact chromatin environment that is refractory to transcriptional activation.

The inhibition of chromatin remodeling, repression of *in vitro* transcription, and polynucleosome compaction activities require only the PSC subunit of the fly PSC; PSC alone has activity levels that are similar to that of the reconstituted complex[54]. Through examining the phenotypes of a series of *Psc* mutants that have different extents of C-terminal truncations in PSC, and the activities of the PSC variants the alleles encode, the functional domain in PSC was mapped to a large region near its C-terminal that has no predicted structure or known homology to other proteins[55]. The analysis of the fly mutants, whose severity of phenotype increases with the extent of truncation in PSC, which in turn correlated with increased loss of *in vitro* activity, provided support for the biological relevance of the three *in vitro* activities of PRC1 described above. However, a caveat to this interpretation is that the PSC truncations also could affect the complex non-specifically, such as compromising the stability of the PSC protein. Thus, the fly phenotypes cannot be specifically linked to the loss of an active domain in PSC.

Reconstituted mouse PRC1, like its fly counterpart, is also able to prevent ATPase-dependent chromatin remodeling and compact polynucleosomes[56]. The conservation of the activities across species provided further support for their biological relevance. The active

component of the mouse complex, however, is CBX2 (previously known as M33), which is the homolog to the fly PC. The murine homologs to PSC, which are Bmi-1 and Mel-18, are inactive in these respects. The active region within CBX2, called 'compaction domain' for its ability to compact polynucleosomes, was mapped and compared to the compaction domain of PSC. Both regions appear to be intrinsically disordered and have an over-representation of the basic residues lysine and arginine. This suggested that an extended region with high positive charge is required for the activities.

Grau *et. al* tested the charge hypothesis by creating a series of CBX2 charge variants and assaying for their *in vitro* activity levels[56]. CBX2 variants of reduced positive charge had different numbers of lysine-to-alanine and arginine-to-alanine amino acid substitutions in the compaction domain, while CBX2 variants of increased positive charge had aspartic acid-to-alanine and glutamic acid-to-alanine substitutions in the same region. The mutations did not prevent the CBX2 variants from being incorporated into PRC1. Grau *et. al* observed an excellent correlation between charge and activity levels of the CBX2 variant, whether in isolation or when incorporated into PRC1. More importantly, they found that the charge requirement is conserved across PcG proteins from other organisms, which is an indicator that this feature of PcG proteins could be relevant for their function *in vivo.* An independent study reported a similar finding[63]. Nonetheless, mutant alleles that specifically affect charge in PSC have not been isolated in flies, presumably because the probability of simultaneously mutating multiple lysine or arginine residues by chance is very small.

Despite the seemingly straightforward connection between polynucleosome compaction, which correlates well with the repression of *in vitro* transcription and chromatin remodeling, to gene repression, its biological relevance remained untested. This is because the nature of an equivalent compact chromatin state in the nucleus is unknown,

and there is no obvious assay for it. Implicit to this model for repression is that DNA becomes inaccessible to transcriptional activators when it is under repression by PRC1. There had been attempts to test for reduced accessibility *in vivo* in flies, but the interpretation of the results was not straightforward.

The approaches to measuring accessibility of DNA typically involve measuring the activity of a heterologous DNA binding protein ('probe') on a region of interest as readout. The approaches fall into two broad categories: the first measures the accessibility of endogenous sequences by treating isolated nuclei with the probes, while the second measures the accessibility of transgenes in the genome using exogenously expressed probes *in vivo.* The assays aimed to determine whether there are differences in the measured accessibility in PcG target genes or PcG regulated regions in a manner that is dependent on the PcG proteins. This was done by comparing the accessibility of a same region between wild type and PcG mutants, or between two cell types where the gene of interest is expressed in one and repressed in the other.

An example of the first type of assay is the study by Schlossherr *et. al* [57]. In this study, the authors used DNA cutting by a series of restriction endonucleases (RE) as the output for accessibility at the *Antp* and *Abd-B* genes. They observed that *Antp* was more accessible in wing discs where it is expressed, compared to the eye-antennal discs, where it was repressed. The authors, however, noted that the magnitude of the differences was small and they were unable to determine if the increased accessibility in the wing discs is a reflection of its actively transcribed state. Moreover, they did not observed a difference in the accessibility of the *Abd-B* gene between wild type and *Pc* mutant embryos. Hence, it is unclear whether accessibility as measured by RE activity is informative of a PcG-mediated function.

Several studies adopted the second type of approach to determine DNA accessibility, that is to measure the accessibility of transgene sequences that are inserted into known PcG-regulated regions using exogenously expressed probes[58-60]. In flies, the transgene is typically inserted into BX-C. These studies collectively showed that there is reduced accessibility that is associated with PcG repression and is dependent on PcG proteins. The magnitude of change is two-fold or less, which is consistent with the report by Schlossherr *et. al* in their study with endogenous PRC1 targets[58]. It was shown that altering transcription alone does not lead to changes in the accessibility. Hence, these results suggest that PcG protein binding and repression is accompanied by reduced access to the target DNA. However, the accessibility changes appear to be sensitive to the characteristic of the transgene insert. The site of the insertion, the number of binding sites for the probe on the transgene, and the distance between the probe binding sites and the region whose accessibility is being measured all appear to influence the accessibility measurement. Hence, it is unclear how the transgene behaviors are representative of endogenous PcG targets, which are diverse in their genomic features such as domain size and nucleotide composition.

The earlier studies measured the accessibility of single loci. More recently, two groups attempted to provide a genome-wide view of the PcG regulated regions[61,62]. Bell *et. al* utilized DNA methylation by the enzyme M. SssI as readout for accessibility. They observed that the average M. SssI methylation level in H3K27me3-decorated regions is lower compared to regions without H3K27me3. The difference of the means, again, is small and there is a large overlap in the range of measured accessibilities between regions with and without H3K27me3. Therefore, it appears that PcG associated regions as a whole have lower accessibility in the genome but the accessibility at individual sites is highly contextual.

In the second study, Deal *et. al* determined the genome-wide nucleosome turnover rates[62]. Heterochromatic regions, which are stably repressed, have low nucleosome turnover rates; thus, it was predicted that PcG repressed regions would have lower turnover rates as well. Surprisingly, the nucleosome in PcG binding sites, together with trxG binding sites, promoters, active gene bodies and replication of origins, have turnover rates that are higher than the genome average. This challenged the conventional view that PcG repressed regions are inert like the heterochromatic regions.

Therefore, there is no strong evidence from chromatin studies that support or refute the model for gene repression through chromatin compaction by PRC1. Given the technical challenges of assaying for a compact chromatin state at the molecular level, we adopted a genetic approach to test the model *in vivo.* An overview of this work is provided in the next section.

DISSERTATION OVERVIEW

The overarching hypothesis of our study is that the *in vitro* activity of polynucleosome compaction is a mechanism by which PRC1 performs its function as an epigenetic transcriptional repressor *in vivo.* We tested the hypothesis via two different approaches, presented in Chapters 2 and 3, respectively.

In Chapter 2, we tested whether the compaction activity is required for the transcriptional repression of PRC1 target genes in cells. We generated mESC lines that express CBX2 variants whose *in vitro* activities were characterized by Grau *et. al*, and assayed for gene expression changes[56]. We observed that the inactive CBX2 variant failed to repress gene expression, while an active CBX2 variant has repressive activities that are similar to wild type. We showed that the inactive CBX2 variant is able to form PRC1 and localize appropriately at PRC1 target genes on the chromatin. Thus, we conclude that the

loss of repressive function is due to the loss of compaction activity in this CBX2 variant. Hence, the compaction activity is required for the transcriptional repression of PRC1 target genes in cells.

In Chapter 3, we tested whether the compaction domain of CBX2 is required for PRC1 function during development. The classic PcG function is defined by its role in maintaining *Hox* gene repression through embryogenesis, which is important for the patterning of the anterior-posterior axis. Hence, we generated mouse mutants with two different knock-in alleles of *Cbx2* with mutations in the compaction domain, and examined their axial skeletons for anterior-posterior transformations. These mice exhibit posterior transformation phenotypes that mimic the phenotypes of *Cbx2*$^{-/-}$ mice, indicating that our *Cbx2* mutant alleles are loss-of-function alleles. Hence, we conclude that the compaction domain is required for PRC1 function during development.

Overall, our studies indicate that the compaction activity is integral to PRC1 function *in vivo.* In Chapter 4, we discuss the implications of these results and suggest some future research directions for understanding the roles of the PcG proteins and chromatin in heritable gene repression.

REFERENCES

1. Sparmann, A. & van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. Nature Reviews Cancer 6, 846–856 (2006).

2. G Struhl, M. A. Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila. *EMBO J* **4,** 3259 (1985).

3. Jürgens, G. A group of genes controlling the spatial expression of the bithorax complex in Drosophila. *Nature* **316,** 153–155 (1985).

4. B, L. E. A gene complex controlling segmentation in Drosophila. *Nature* **276,** 565–570 (1978).

5. Durston, A. J. Hox Genes: Master Regulators of the Animal Bodyplan, Embryogenesis, (Ed.), ISBN: 978-953-51-0466-7, InTech

6. Driever, W. & Nüsslein-Volhard, C. A gradient of bicoid protein in Drosophila embryos. *Cell* **54,** 83–93 (1988).

7. Driever, W. & Nüsslein-Volhard, C. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell* **54,** 95–104 (1988).

8. Lewis. E. B New mutants: Reports of P. Lewis. . *Drosoph. Inf. Serv.* **21,** 69 (1947).

9. Moazed, D. & O'Farrell, P. H. Maintenance of the engrailed expression pattern by Polycomb group genes in Drosophila. *Development* **116**, 805-810 (1992).

10. Pearce, J. J., Singh, P. B. & Gaunt, S. J. The mouse has a Polycomb-like chromobox gene. *Development* **114,** 921–929 (1992).

11. van der Lugt, N. M. *et al.* Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes and Development* **8**, 757-769 (1994).

12. Akasaka, T. *et al.* A role for mel-18, a Polycomb group-related vertebrate gene, during theanteroposterior specification of the axial skeleton. *Development* **122,** 1513–1522 (1996).

13. Takihara, Y. *et al.* Targeted disruption of the mouse homologue of the Drosophila polyhomeotic gene leads to altered anteroposterior patterning and neural crest defects. *Development* **124,** 3673–3682 (1997).

14. Muller, J., Gaunt, S. & Lawrence, P. A. Function of the Polycomb protein is conserved in mice and flies. *Development* **121,** 2847–2852 (1995).

15. Gao, Z. *et al.* PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Molecular cell* **45,** 344–356 (2012).

16. Aranda, S., Mas, G. & Di Croce, L. Regulation of gene transcription by Polycomb

proteins. *Sci Adv* **1,** e1500737 (2015).

17. Senthilkumar, R. & Mishra, R. K. Novel motifs distinguish multiple homologues of Polycomb in vertebrates: expansion and diversification of the epigenetic toolkit. *BMC Genomics* **10,** 549 (2009).

18. Ma, R.-G., Zhang, Y., Sun, T.-T. & Cheng, B. Epigenetic regulation by polycomb group complexes: focus on roles of CBX proteins. *J Zhejiang Univ Sci B* **15,** 412–428 (2014).

19. Cbx4<tm1.2Gxu> Targeted Allele Detail MGI Mouse (MGI:5489932). *informatics.jax.org*

20. Cbx8<tm1.1Hko> Targeted Allele Detail MGI Mouse (MGI:5306630). *informatics.jax.org*

21. Morey, L. *et al.* Nonoverlapping Functions of the Polycomb Group Cbx Family of Proteins in Embryonic Stem Cells. *Cell Stem Cell* **10,** 47–62 (2012).

22. Creppe, C., Palau, A., Malinverni, R., Valero, V. & Buschbeck, M. A Cbx8-Containing Polycomb Complex Facilitates the Transition to Gene Activation during ES Cell Differentiation. *PLoS Genet.* **10,** e1004851 (2014).

23. Sharp, E. J., Martin, E. C. & Adler, P. N. Directed overexpression of suppressor 2 of zeste and Posterior Sex Combs results in bristle abnormalities in Drosophila melanogaster. *Developmental biology* **161,** 379–392 (1994).

24. Rastelli, L., Chan, C. S. & Pirrotta, V. Related chromosome binding sites for zeste, suppressors of zeste and Polycomb group proteins in Drosophila and their dependence on Enhancer of zeste function. *EMBO J* **12,** 1513–1522 (1993).

25. Paro, R. & Hogness, D. S. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proc. Natl. Acad. Sci. U.S.A.* **88,** 263–267 (1991).

26. Wang, L. *et al.* Hierarchical Recruitment of Polycomb Group Silencing Complexes. *Molecular Cell* **14,** 637–646 (2004).

27. Schwartz, Y. B. *et al.* Genome-wide analysis of Polycomb targets in Drosophila melanogaster. *Nat Genet* **38,** 700–705 (2006).

28. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nat Cell Biol* **441,** 349–353 (2006).

29. Kennison, J. A. & Tamkun, J. W. Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.* **85,** 8136–8140 (1988).

30. Pengelly, A. R., Kalb, R., Finkl, K. & Müller, J. Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes & Development* (2015). doi:10.1101/gad.265439.115

31. Müller, J. *et al.* Histone Methyltransferase Activity of a Drosophila Polycomb Group Repressor Complex. *Cell* **111,** 197–208 (2002).

32. Shao, Z. *et al.* Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98,** 37–46 (1999).

33. Francis, N. J., Saurin, A. J., Shao, Z. & Kingston, R. E. Reconstitution of a functional core polycomb repressive complex. *Molecular Cell* **8,** 545–556 (2001).

34. Gutiérrez, L. *et al.* The role of the histone H2A ubiquitinase Sce in Polycomb repression. *Development* **139,** 117-127 (2012).

35. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nat Cell Biol* **431,** 2799–2810 (2004).

36. de Napoles, M. *et al.* Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Developmental Cell* **7,** 663–676 (2004).

37. Hernandez-Munoz, I. *et al.* Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin E3 ligase. *Proceedings of the National Academy of Sciences* **102,** 7635–7640 (2005).

38. Fang, J., Chen, T., Chadwick, B., Li, E. & Zhang, Y. Ring1b-mediated H2A Ubiquitination Associates with Inactive X Chromosomes and Is Involved in Initiation of X Inactivation. *Journal of Biological Chemistry* **279,** 52812–52815 (2004).

39. Lagarou, A. *et al.* dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes & Development* **22,** 2799–2810 (2008).

40. Stock, J. K. *et al.* Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* **9,** 1428–1435 (2007).

41. Blackledge, N. P. *et al.* Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell* **157,** 1445-1459 (2014).

42. Cooper, S. *et al.* Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. *Cell Reports* (2014). doi:10.1016/j.celrep.2014.04.012

43. Tavares, L., Dimitrova, E., Oxley, D., Webster, J. & Poot, R. RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3. *Cell* (2012). doi:10.1016/j.cell.2011.12.029

44. Cao, R., Tsukada, Y.-I. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A Ubiquitylation and Hox Gene Silencing. *Molecular Cell* **20,** 845–854 (2005).

45. Morey, L., Aloia, L., Cozzuto, L., Benitah, S. A. & Di Croce, L. RYBP and Cbx7 Define Specific Biological Functions of Polycomb Complexes in Mouse Embryonic Stem Cells. *Cell Reports* **3,** 60–69 (2013).

46. Illingworth, R. S. *et al.* The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. *Genes & Development* **29,** 1897–1902 (2015).

47. Buchwald, G. *et al.* Structure and E3-ligase activity of the Ring–Ring complex of Polycomb proteins Bmi1 and Ring1b. *EMBO J* **25,** 2465–2474 (2006).

48. Kim, C. A., Gingery, M., Pilpa, R. M. & Bowie, J. U. The SAM domain of polyhomeotic forms a helical polymer. *Nat Struct Mol Biol* **9,** 453–457 (2002).

49. Lavigne, M., Francis, N. J., King, I. F. G. & Kingston, R. E. Propagation of Silencing. *Molecular Cell* **13,** 415–425 (2004).

50. Isono, K. *et al.* SAM Domain Polymerization Links Subnuclear Clustering of PRC1 to Gene Silencing. *Developmental Cell* **26,** 565–577 (2013).

51. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* (2016). doi:10.1038/nature16496

52. Wani, A. H. *et al.* Chromatin topology is coupled to Polycomb group protein subnuclear organization. *Nat Comms* **7,** 10291 (2016).

53. King, I. F. G., Francis, N. J. & Kingston, R. E. Native and Recombinant Polycomb Group Complexes Establish a Selective Block to Template Accessibility To Repress Transcription In Vitro. *Mol. Cell. Biol.* **22,** 7919-7928 (2002).

54. Francis, N. J. Chromatin Compaction by a Polycomb Group Protein Complex. *Science* **306,** 1574–1577 (2004).

55. King, I. F. G. *et al.* Analysis of a Polycomb Group Protein Defines Regions That Link Repressive Activity on Nucleosomal Templates to In Vivo Function. *Mol. Cell. Biol.* **25,** 6578–6591 (2005).

56. Grau, D. J. *et al.* Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes & Development* **25,** 2210–2221 (2011).

57. Schloßherr, J., Eggert, H., Paro, R., Cremer, S. & Jack, R. S. Gene inactivation in Drosophiia mediated by the Polycomb gene product or by position-effect variegation does not involve major changes in the accessibility of the chromatin fibre. *Molec. Gen. Genet.* **243,** 453–462 (1994).

58. Boivin, A. & Dura, J.-M. In vivo chromatin accessibility correlates with gene silencing in Drosophila. *Genetics* **150,** 1539-1549 (1998).

59. McCall, K. & Bender, W. Probes of chromatin accessibility in the Drosophila bithorax complex respond differently to Polycomb-mediated repression. *EMBO J* **15,** 569

(1996).

60. Fitzgerald, D. P. & Bender, W. Polycomb group repression reduces DNA accessibility. *Mol. Cell. Biol.* **21,** 6585–6597 (2001).

61. Bell, O. *et al.* Accessibility of the Drosophila genome discriminates PcG repression, H4K16 acetylation and replication timing. *Nat Struct Mol Biol* **17,** 894–900 (2010).

62. Deal, R. B., Henikoff, J. G. & Henikoff, S. Genome-Wide Kinetics of Nucleosome Turnover Determined by Metabolic Labeling of Histones. *Science* **328,** 1161–1164 (2010).

63. Beh, L. Y., Colwell, L. J. & Francis, N. J. A core subunit of Polycomb repressive complex 1 is broadly conserved in function but not primary sequence. *Proc. Natl. Acad. Sci. U.S.A* E1063-1071 (2012).

CHAPTER 2

CBX2 with compaction activity is required for gene repression in cells

Contributions:

Sharmistha Kundu (Kingston lab) generated the ChIP-seq data in Figure 2.1B. John Seavitt, (MGH Gene Targeting Facility) provided technical help in generating the doxycycline-inducible mouse embryonic stem cell lines. Peggy Wang and Ruslan Sadreyev (Molbio Bioinformatics Core) performed the ChIP-seq and RNA-seq analyses in Figures 2.2 and 2.3. Andrej Savol (Molbio Bioinformatics Core) generated Figure 2.2E. Sharmistha Kundu performed the ChIP experiments in Figure 2.5.

ABSTRACT

Polycomb group (PcG) proteins are developmental regulators that maintain their target genes in the repressive state. The molecular mechanisms by which they achieve repression remain to be elucidated. Polycomb repressive complex 1 (PRC1) is able to compact polynucleosomes *in vitro,* but the relevance of this activity to its function *in vivo* is unknown. For mammalian PRC1, the activity is mapped to a basic region in its CBX2 subunit. Mutations that reduce charge in this region disrupt the compaction activity of the complex. We tested the hypothesis that the compaction activity is important for PRC1-mediated gene repression in cells. We compared the gene expression patterns in mouse embryonic stem cell lines expressing CBX2 variants whose *in vitro* activities were characterized. We observed that the inactive CBX2 variant failed to repress gene expression, while an active CBX2 variant has repressive activities that are similar to wild type. We showed that the inactive CBX2 variant is able to form PRC1 and localize appropriately at PRC1 target genes on the chromatin. In addition, we found that another CBX homolog CBX7 and the Polycomb –associated histone modifications H3K27me3 and H2AK119Ub are insufficient for full repression at some genes in the absence of CBX2. Thus, we conclude that the loss of repressive function is due to the loss of compaction activity in the inactive CBX2 variant. Our results support the hypothesis that chromatin compaction is relevant to PRC1 function *in vivo.*

INTRODUCTION

The *Polycomb* group (PcG) genes were first identified through genetic studies in flies. They were identified because they produced an interesting mutant phenotype where a particular body segment assumed the identity of another, i.e. homeotic transformation. The PcG genes maintain the repression of *Hox* genes through embryogenesis for proper patterning of the anterior-posterior axis. In addition, they regulate many other developmental genes. The PcG genes and the *Hox* genes are widely conserved from fly through mammals. Mutations in the PcG genes in mice also result in homeotic transformation phenotypes that are analogous to those observed in the fly mutants, suggesting that the gene regulatory roles of PcG genes are conserved.

Several lines of evidence indicated the PcG proteins interact with chromatin. The fly PC protein has a chromodomain that was also present in Heterochromatin protein 1 (HP1), which associates with heterochromatin[1]. Secondly, mutations in *trithorax* group (trxG) genes were found to suppress PcG phenotypes[2]. Several of the trxG genes encode ATP-dependent chromatin remodelers that move nucleosomes to maintain active transcription; this indicated PcG gene products may antagonize chromatin remodeling. Moreover, several PcG proteins, e.g. PC, PSC and PH were observed to bind to polytene chromosomes at sites that included the *Hox* gene clusters ANT-C and BX-C[3,4].

In order to gain mechanistic understanding of PcG-mediated gene repression, PcG proteins have been purified and subjected to assays that investigate their interactions with the nucleosome or a chromatin template. The proteins form two broad classes of complexes: PRC1 and PRC2. This study focuses on PRC1. The core components of the fly PRC1 are Polycomb (PC), Polyhomeotic (PH), Posterior sex combs (PSC) and Sex comb extra

(SCE). PRC1 and its core components are conserved in mammals but with added complexity; there are two to six homologs for each of the components in mammals (Figure 1.2). The homologs can substitute for one another for PRC1 formation[5]; however, they are not equivalent in their biochemical activities. They also have different but overlapping temporal and spatial expression patterns during development and in the organism. Therefore, PcG-mediated gene repression appears to be much more complex in mammals than in flies.

The biochemical activities of each core PRC1 component have been characterized and they form the basis for models of how PRC1 maintain gene repression. SCE, when in complex with PSC, mono-ubiquitylates histone H2A (H2AUb)[6,7]. This activity is conserved in specific mammalian Ring1/PCGF homolog combinations. H2AUb was associated with RNA polymerase pausing and the recruitment of PRC2 for gene repression in cells[8] However, recent studies with SCE and Ring1B catalytic mutants in flies and mice, respectively, indicate that it is unlikely to be directly involved in PRC1-dependent repression[9,10]. PH and all three mammalian Phc proteins are able to self-assemble into oligomers via their SAM domains. This activity was demonstrated to be important for higher order chromatin folding and PRC1 gene repression in mice and in fly cells[11,12]. PC and the CBX homologs contain a chromodomain that binds to the histone modification H3K27me3. They are therefore thought to play a role in targeting PRC1 to genomic regions delineated by H3K27me3, which is shown to be important for gene repression[13]. The CBX proteins are also able to bind H3K9me. Different CBX homologs show varying affinities for either histone mark that can be modulated by post-translational modifications on the protein itself [14-16].

An *in vitro* reconstituted fly PRC1 containing the core components without PH has the ability to inhibit chromatin remodeling and to create compact structures on nucleosomal arrays[17,18]. These two activities correlate very well, and this gave rise to a model where PRC1 creates a localized chromatin structure that is refractory to chromatin remodeling and transcription in the nucleus. However, general DNA accessibility is not markedly reduced at PcG-repressed regions *in vivo*[19-21].

The PSC subunit alone is sufficient for the two activities[18]. The functional domain in PSC is a large region near its C-terminus that has no predicted structure or known homology to other proteins[22]. A study of PSC truncation mutants hinted at the biological relevance of the two *in vitro* activities. The severity of their phenotypes correlated with the extent of the truncations from their C-termini, suggesting that an increased loss of the functional domain contributed to concomitant loss of PRC1 function.

Another indication that the *in vitro* activities are relevant is that they are conserved in the *in vitro* reconstituted mouse PRC1[23]. The activities, however, are found in the homolog of PC, CBX2, instead of the homolog to PSC, which is Bmi-1. The functional domain in CBX2 was identified, and like that for PSC, it has no predicted structure or known homology domains. A comparison of the two functional domains revealed a bias in amino acid composition for the basic residues lysine (K) and arginine (R). These gave the two proteins high overall positive charge. To test whether the K and R residues are important, Grau *et. al* delineated two regions in the functional domain and mutated every other K and R residue to A (K-to-A, and R-to-A). They found the mutations abrogated the activities, and partial mutations led to partial loss of activity. In contrast, aspartic acid-to-alanine (D-to-A) and glutamic acid-to-alanine (E-to-A) mutations that increase charge did not affect the activities. Therefore, there appears to be a correlation between charge in CBX2 and its

activities. However, positive charge is not the sole requirement, as non-PcG proteins of similar charge to CBX2 did not have such activities. Lastly, Grau *et. al* found that high overall positive charge is a good predictor for activity in putative PRC1 components in other organisms, including zebrafish, frog and worm. This has two implications: firstly, the chromatin remodeling inhibition and chromatin compaction activities are widely conserved; secondly, the molecular basis for the activities is also likely to be conserved. An independent study by Beh *et. a*l reported a similar finding[24].

Therefore, we hypothesize that K and R residues in CBX2 that confer it its compaction activities are important for PcG gene repression in cells. We test the hypothesis by comparing gene expression changes in cell lines that express wild type or CBX2 charge mutants previously characterized by Grau *et. al.*

RESULTS

CBX2 is expressed and binds to PRC1 target genes in mESCs

We needed a cell culture system to investigate whether the K and R residues in the compaction domain of CBX2 are required for repression of PRC1 target genes. There are five CBX homologs that are expressed in a cell-type specific manner, namely CBX2, 4, 6, 7 and 8[25]. In mouse embryonic stem cells (mESCs), CBX2, 6 and 7 are expressed. However, only CBX2 and 7 are associated with PRC1. The expression level of CBX2 is low compared to that of CBX7 in mESCs. CBX2 depletion does not affect mESC self-renewal, unlike CBX7 which is important for both self-renewal and pluripotency[25,26]. Given that *Cbx2*$^{-/-}$ mESCs are

stable in culture, we reasoned that we could express CBX2 variants in them to assess their effects on gene expression.

Earlier studies have shown that CBX2 is expressed and binds to known Polycomb target genes in mESCs, where Ring1B and the PRC2 components Eed and Suz12 also bind[27,28]. We wanted to verify this to ensure that CBX2 may have PRC1-associated roles in mESCs. To confirm that CBX2 is present in wild type mESCs, we analyzed four common mESC lines by western blot analysis. We detected a band for CBX2 in all four wild type lines but not in the *Cbx2*$^{-/-}$ mESC line, which we used as a negative control for the specificity of the antibody (Figure 2.1A). To verify that CBX2 binds to chromatin in the context of PRC1, we performed ChIP-seq analyses on mESCs for CBX2, Ring1B, Phc1 and CBX7. The CBX2 binding sites coincide with those for the other three proteins (Figure 2.1B). The antibody used for ChIP did not give enrichment signal in *Cbx2*$^{-/-}$ mESCs (see Figure 2.2D), indicating that the antibody used for CBX2-ChIP is specific for CBX2.

These data indicate that CBX2 is expressed in mESCs and forms PRC1 that localizes onto chromatin. Hence, we conclude that mESCs are an appropriate system for comparing the effect of wild type (WT) and mutant CBX2 on gene repression. Notably, CBX7 is also present at CBX2 binding sites, indicating that there are at least two types of CBX-containing PRC1 in mESCs.

Generation of cell lines that express WT and mutant CBX2

To generate comparable mESC lines that express either WT or mutant CBX2, we used a Tet-ON system which re-introduces WT-CBX2, KRA-CBX2, or DEA-CBX2 expression in *Cbx2*$^{-/-}$ mESCs upon doxycycline treatment. The KRA-CBX2 has a total of 23 lysine-to-alanine (K-to-A) and arginine-to-alanine (R-to-A) substitutions in two regions that are

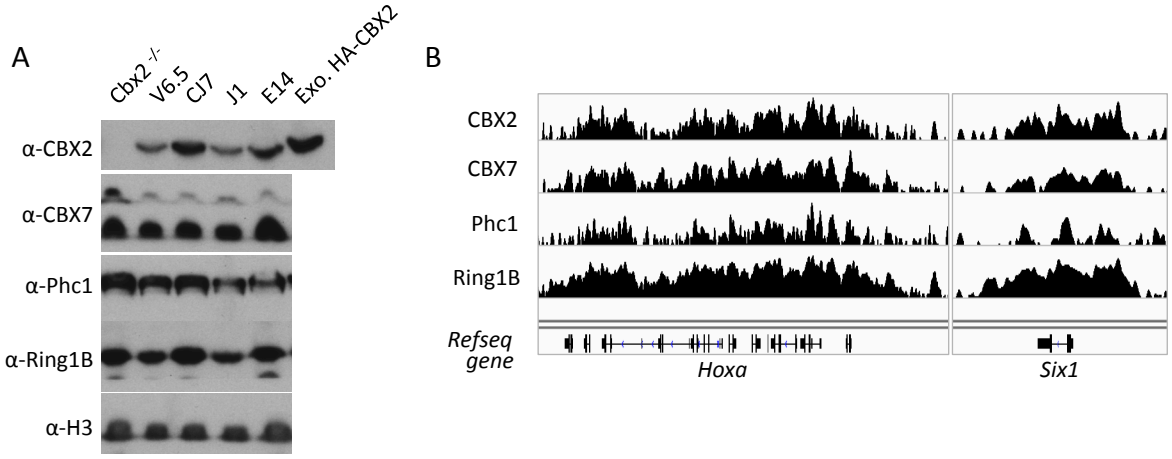


Figure 2.1 CBX2 is expressed and binds to PRC1 target genes in mESCs.
A. Western blot shows the expression of CBX2 in four common mESC lines. Lysate from *Cbx2*$^{-/-}$ mESCs and exogenous HA-tagged CBX2 were used as negative and positive controls, respectively, for the specificity of the CBX2 antibody. This panel also shows the presence of other PRC1 components, CBX7, Phc1 and Ring1B, in these cells. H3 is used as loading control for the lysates. **B.** Genome browser tracks of ChIP-seq data for CBX2 and other PRC1 subunits at two known PRC1 target regions.

required for its *in vitro* activities (Figure 2.2A). The mutations reduce the positive charge on the protein. In contrast, the DEA-CBX2 has aspartic acid-to-alanine (D-to-A) and glutamic acid-to-alanine (E-to-A) substitutions in the same regions, which increase the positive charge. Both KRA-CBX2 and DEA-CBX2 were previously characterized biochemically[23]: compared to WT-CBX2, KRA-CBX2 was unable to inhibit chromatin remodeling and compact nucleosomal arrays, while DEA-CBX2 has activity levels similar to WT-CBX2.

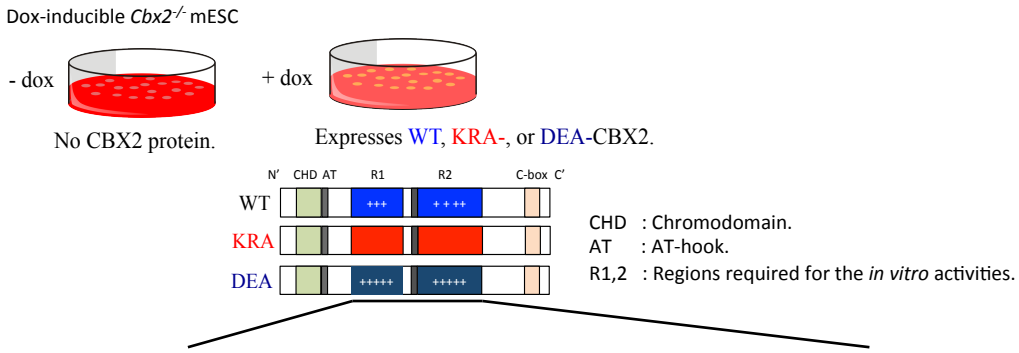
We were able to determine the conditions that are required for the mESC lines to express CBX2 to similar levels (Figure 2.2B). We did this for two independent WT-CBX2 lines, two independent KRA-CBX2 lines and one DEA-CBX2 line, which were all used in our gene expression studies reported below. The presence of mutant CBX2 proteins does not affect the protein levels of other PRC1 components, as shown for CBX7 and Phc1 by western blot analysis (Figure 2.2B).

Complex formation and targeting of KRA- and WT-CBX2

To test whether KRA-CBX2 is incorporated into PRC1 complex in the cells, we immunoprecipitated Ring1B and assayed for CBX2 by western blot analysis. KRA-CBX2 interacts with Ring1B, as does the re-introduced WT-CBX2 (Figure 2.2C). Thus, we conclude that KRA-CBX2 can interact with other PRC1 subunits for complex assembly. This is consistent with previous findings that the amino acid substitutions do not affect CBX2-Ring1B interaction, which occurs through its C-box domain[23].

To determine whether KRA-CBX2 localizes to *bona fide* PRC1 target genes, we performed ChIP-seq analysis for KRA-CBX2 and compared it to that for WT-CBX2. We observed that the ChIP enrichment signals for KRA-CBX2 coincide to a high degree with those for WT-CBX2 at known PRC1 target genes (Figure 2.2D): 263 out of the 285 genes

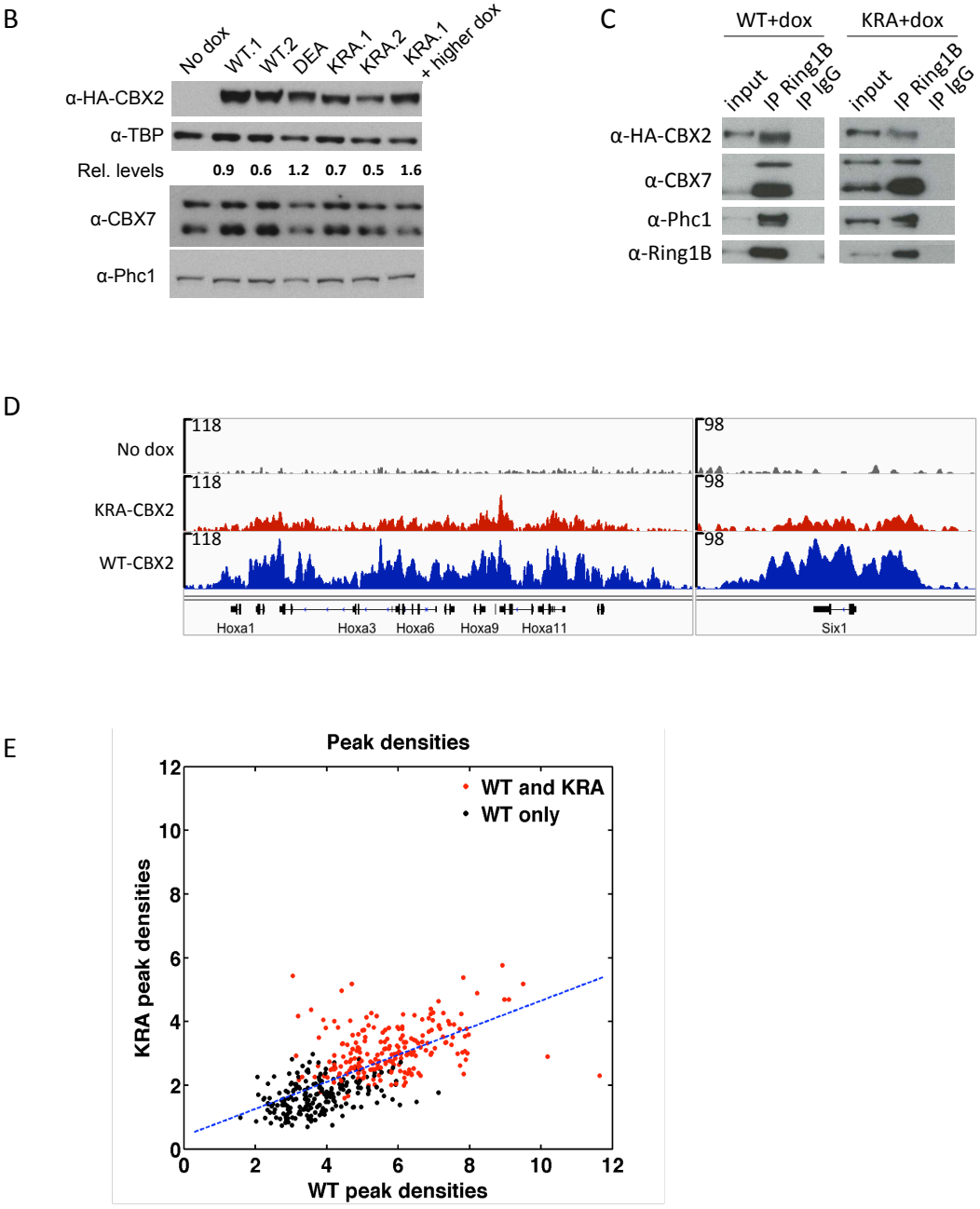
A




```
WT-CBX2          DEEEDDSDLDSKRGPRGRETHPVPQKKAQILVAKPELKDPIRKKRGRKPLPPEQKAARRP 180
DEA-CBX2         DEEEDDSDLDSKRGPRGRETHPVPQKKAQILVAKPALKDPIRKKRGRKPLPPAQKAARRP 180
KRA-CBX2         DEEEDDSDLDSARGPAGRETHPVPQAKAQILVAAPELKDPIAKARGAKPLPPEQAAARAP 180
                 *********** *** ********* ******* * ***** * ** ***** * *** *

WT-CBX2          VSLAKVLKTTRKDLGTSAAKLPPPLSAPVAGLAALKAHTKEACGGPSTMATPENLASLMK 240
DEA-CBX2         VSLAKVLKTTRKDLGTSAAKLPPPLSAPVAGLAALKAHTKAACGGPSTMATPENLASLMK 240
KRA-CBX2         VSLAKVLATTRADLGTSAAKLPPPLSAPVAGLAALAAHTKEACGGPSTMATPENLASLMA 240
                 ******* *** ********************** **** ******************

WT-CBX2          GMAGSPSRGGIWQSSIVHYMNRMSQSQVQAASRLALKAQATNKCGLGLDLKVRTQKGGEL 300
DEA-CBX2         GMAGSPSRGGIWQSSIVHYMNRMSQSQVQAASRLALKAQATNKCGLGLALKVRTQKGGEL 300
KRA-CBX2         GMAGSPSRGGIWQSSIVHYMNAMSQSQVQAASRLALAAQATNKCGLGLDLAVRTQAGGEL 300
                 ******************** ************** *********** * **** ****

WT-CBX2          GGSPAGGKVPKAPGGGAAEQQRGNHSGSPGAQLAPTQELSLQVLDLQSVKNGVPGVGLLA 360
DEA-CBX2         GGSPAGGKVPKAPGGGAAAQQRGNHSGSPGAQLAPTQELSLQVLALQSVKNGVPGVGLLA 360
KRA-CBX2         GGSPAGGKVPAAPGGGAAEQQRGNHSGSPGAQLAPTQELSLQVLDLQSVANGVPGVGLLA 360
                 ********** ******* ************************* **** **********

WT-CBX2          RHAPAKAIPATNPATGKGPGSGPTGANMTNAPTDNNKGEKLTCKATALPAPSVKRDTVKS 420
DEA-CBX2         RHAPAKAIPATNPATGKGPGSGPTGANMTNAPTDNNKGAKLTCKATALPAPSVKRDTVKS 420
KRA-CBX2         RHAPAAAIPATNPATGKGPGSGPTGANMTNAPTDNNAGEKLTCAATALPAPSVKADTVKS 420
                 ***** ****************************** * **** ********** *****
```

Figure 2.2 WT- and KRA-CBX2 are expressed similarly, interact with Ring1B and bind to CBX2 target genes. A. A schematic illustration of the doxycycline inducible system used to express WT-, KRA- and DEA-CBX2 in *Cbx2*$^{-/-}$ mESCs. The amino acid sequence for regions 1 and 2 (R1, R2) of the CBX2 variants is shown. The substituted amino acid residues are highlighted in blue and red. **B.** Western blot showing the CBX2 levels in the WT-, KRA- and DEA-CBX2 mESC lines used in subsequent experiments. The protein levels are determined via ImageJ using TBP levels as the loading control. The values are the average measurement from two technical replicates and presented as relative values to the level in the WT.1 line. The levels of CBX7 and Phc1 are also shown. **C.** Western blots showing co-immunoprecipitation of WT- and KRA-CBX2 with Ring1B. CBX7 and Phc1 also co-immunoprecipitated with Ring1B as expected. **D.** Genome browser tracks of ChIP-seq data for CBX2 from WT- and KRA-CBX2 expressing cells at known PRC1 target regions. **E.** Plots of ChIP signal densities for KRA-CBX2 against those for WT-CBX2 at every genomic region where a peak is called from the WT-CBX2 ChIP-seq data set. In red are the regions where a peak is also called from the KRA-CBX2 ChIP-seq data. In black are the regions that are not called as peaks from the KRA-CB2 ChIP-seq data.

Figure 2.2 (Continued)



with KRA-CBX2 binding are WT-CBX2 bound genes (92% overlap). Of the WT-CBX2 bound genes, 78% of them are genes with endogenous CBX2 binding, as determined from ChIP-seq experiments with WT mESCs. Therefore, we concluded that KRA-CBX2, like the reintroduced WT-CBX2, is able to localize appropriately to PRC1 target genes.

We noted that the ChIP signal for KRA-CBX2 is weaker compared to that for WT-CBX2. Given that crosslinking between protein and DNA occurs via the primary amine group on lysine residues, we expect KRA-CBX2 to crosslink less efficiently than the WT protein due to the K-to-A substitutions introduced. To test whether the ChIP enrichment signals for KRA-CBX2 are overall lower than those for WT-CBX2, we plotted the peak densities for KRA against the peak densities for WT for every peak that is called from the WT-CBX2 ChIP-seq data set. The two correlated linearly, and the best-fit line has a slope of 0.4 (Figure 2.2E). This indicates that the KRA peaks densities are generally about half of those of the WT peaks. We later observed the same with our ChIP-qPCR experiments (see Figure 2.5A). We observed that the peaks that are called from the KRA-CBX2 ChIP-seq data set are mainly the peaks with high densities in the WT-CBX2 data set (Figure 2.2E). This suggests that only genomic regions with high CBX2 binding showed sufficient enrichment above noise in the KRA-CBX2 ChIP-seq experiment to be called as peaks. These observations are consistent with the explanation that poor crosslinking with KRA-CBX2 leads to the weaker overall ChIP enrichment compared to WT. However, we note that these observations do not rule out poorer binding or occupancy efficiencies for KRA-CBX2 compared to WT-CBX2.

We conclude that in our cell lines, the KRA-CBX2 forms PRC1 and binds to PRC1 target genes. Thus, we can assay for the effects of the mutations on transcriptional regulation.

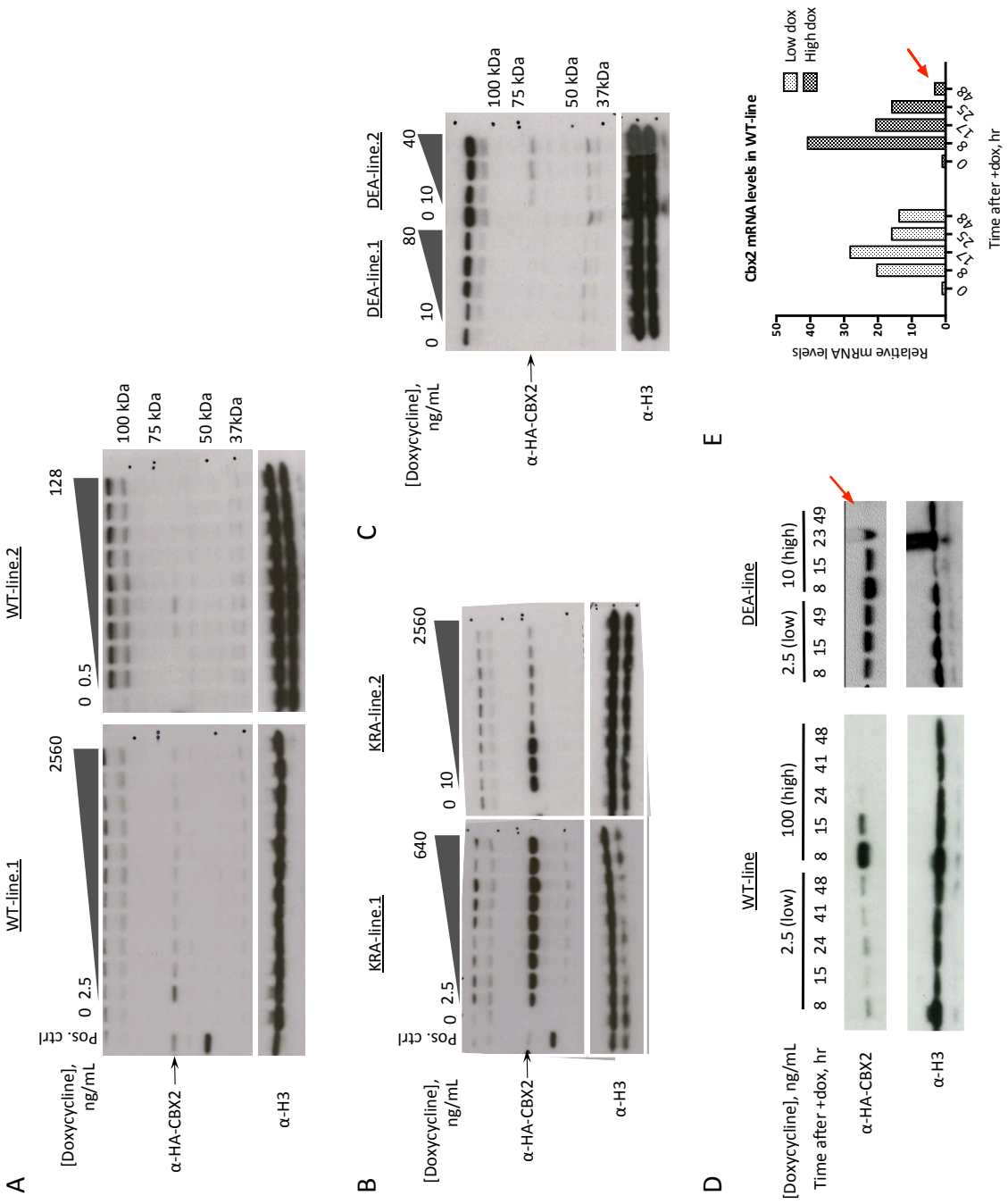
Tight regulation of WT- and DEA-CBX2 levels, but not for KRA-CBX2

We noted an interesting observation while performing the doxycycline titration experiments to determine the conditions required for the cell lines to express CBX2 to similar levels. We observed the KRA-CBX2 lines showed increased CBX2 protein levels with increasing doxycycline as expected, but the WT and DEA-CBX2 lines appeared to fail to respond to increasing doxycycline (Figure 2.3A, B, C). This held true for independently derived mESC clones. We suspected that the WT- and DEA-CBX2 proteins could not be expressed at high levels in the cells because of their repressive activities. Thus, we followed the protein levels after the addition of doxycycline over a time period of 48 hours to determine if there is a burst of expression that is rapidly down regulated. Indeed this was what we observed (Figure 2.3D). When the cells were treated with high concentrations of doxycycline, CBX2 was made but was quickly lost by 48 hours after induction. On the other hand, when the cells were treated with low concentrations of doxycycline, CBX2 levels remained relatively stable. The mRNA levels of *Cbx2* correlated with the protein levels, indicating that protein loss is due to downregulation of transcription of the transgene (Figure 2.3E). The tight regulation of WT-and DEA-CBX2 levels is reminiscent of dosage sensitivity *Polycomb* group mutants in flies and mice, where both haploinsufficiency and overexpression of the *Polycomb* group genes lead to developmental abnormalities[29-31].

These observations indicate that KRA- and WT-CBX2 are different with respect to their tolerable levels in mESCs.

Figure 2.3 The expressions of WT- and DEA-CBX2 , but not KRA- CBX2, are tightly regulated in mESCs. A. to C. Western blots showing CBX2 levels in the cell lines in response to increasing amounts of doxycycline used for induction. Panels **A**, **B** and **C** are the responses for WT-, KRA- and DEA- cell lines, respectively. Lines.1 and .2 are two independent mESC clones for each CBX2 variant line. **D.** Western blots showing CBX2 levels in WT-and DEA- cells in response to low or high doxycycline induction over 48 hours. The protein levels are stable at low doxycycline concentrations but decrease over time at high doxycycline concentrations. The red arrow highlights the complete loss of proteins in the DEA-line at 48 hours. **E.** RT-qPCR results showing mRNA levels of *Cbx2* at low or high doxycycline induction over 48 hours. The red arrow indicates the mRNA level has decreased to near no induction levels. *Actb* mRNA was used as loading control. The levels are relative to the amount before induction. n=1.

Figure 2.3 (Continued)



Effect of charge mutations on gene expression

To investigate the effect of the charge mutations in CBX2 on gene expression, we performed RNA-seq of the cell lines before and after doxycycline treatment that induces CBX2 expression.

We wanted to have set(s) of genes that may be regulated by CBX2 on which we will perform differential expression analyses. To do this we identified genes that have CBX2 binding from our WT-CBX2 and KRA-CBX2 ChIP seq datasets and created three lists: genes with WT-CBX2 binding, genes with KRA-CBX2 binding, and genes that are found in both lists. From these, we selected genes that had detectable transcript levels in at least one sample from our RNA-seq data. The resulting gene lists had 130, 67 and 59 genes, respectively (Figure 2.4 Venn diagram). For these genes we calculated the fold change in mRNA levels as a ratio of post-induction mRNA counts to pre-induction counts (Fold Change = +doxycycline/ - doxycycline). By this metric, genes with a negative fold change values would be genes that become repressed following CBX2 expression, and vice versa. We did this for the five mESC lines shown in Figure 2.2B, and an additional sample where KRA.1 was treated with more doxycycline to express KRA-CBX2 at a higher level. The analyses are presented in heat maps where red represents down regulation and blue for upregulation (Figure 2.4A).

We first focused our analysis on the genes with both WT- and KRA-CBX2 binding. We found that re-introducing WT CBX2 resulted in the repression of many of these genes (Figure 2.4A, left panel, left two columns). The repression is also evident when screen shots of RNA-seq tracks of affected genes were examined (Figure 2.4B, blue tracks). Notably, non-PRC1 target genes do not exhibit this decrease. In comparing the effects of re-introducing

Figure 2.4 The effects of WT-, KRA- and DEA-CBX2 on gene expression changes. A. Heat maps showing changes in gene expression at genes with both WT- and KRA- CBX2 binding (left), all genes with WT-CBX2 binding (center), and all genes with KRA-CBX2 binding (right). Red and blue represent down regulation and up regulation, respectively. Hierarchical clustering was performed on the cell lines according the patterns of gene expression changes. WT.1, WT.2, and KRA.1, KRA.2, are independent cell lines expressing WT- and KRA-CBX2, respectively. **B.** Screenshots of RNAseq tracks showing the read counts, after normalization to library size, of mRNA . The top two panels show two known PRC1 targets that are repressed by WT- and DEA-CBX2, but not by KRA-CBX2. The bottom two panels show that non-PRC1 target genes are not affected by any of the CBX2 variants. The green track is CBX2 ChIP seq enrichment signal over the gene. **C.** Validation of the gene expression changes by RT-qPCR experiments. The values plotted are Mean ± SD, n=4. Statistical significance was determined using t-tests comparing the means of –dox and +dox at individual genes ** $P < 0.01$; * $P < 0.05$. **D.** Matrix for Pearson correlation scores of the cell lines based on their gene expression changes.

Figure 2.4 (Continued)

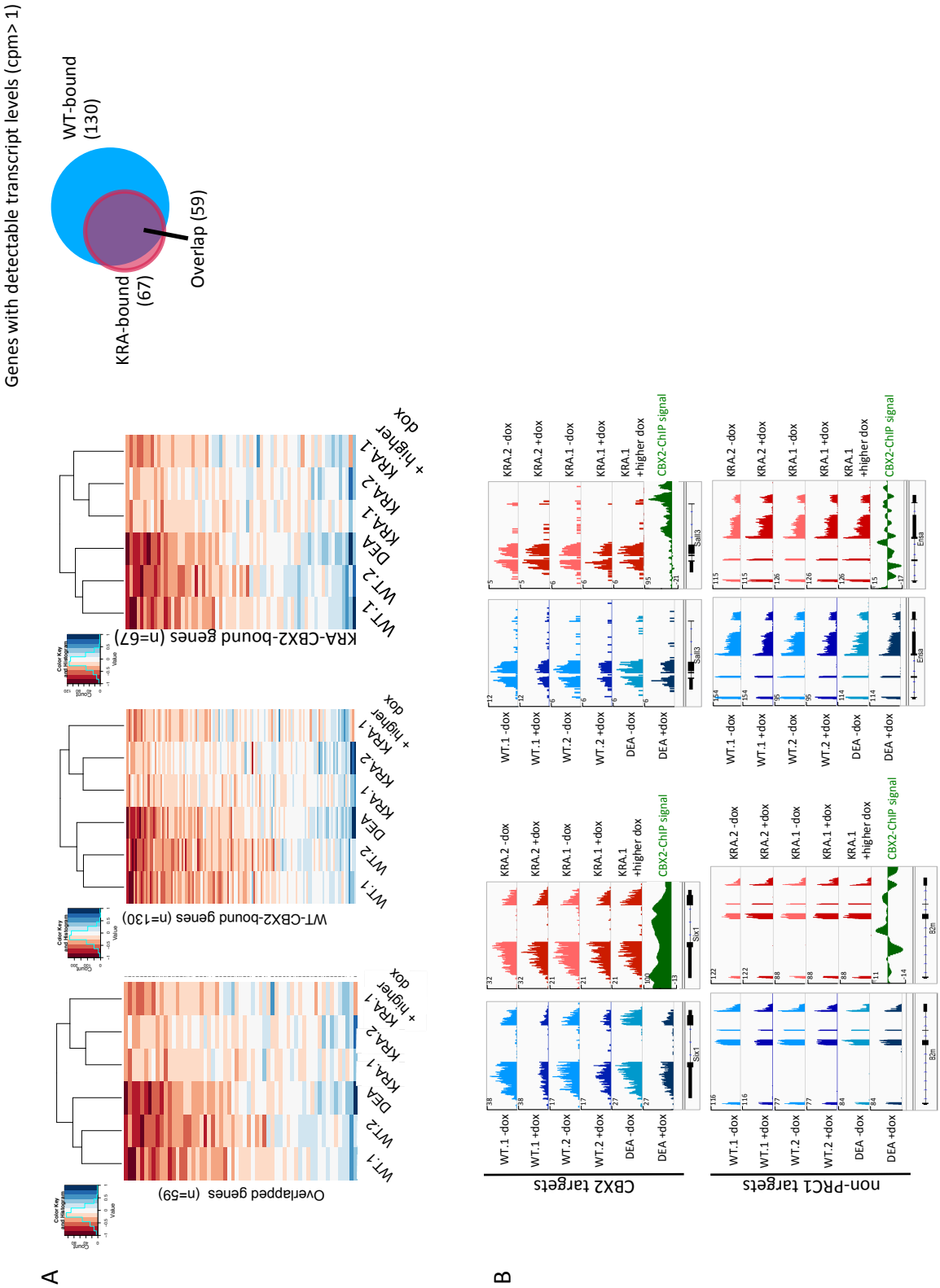
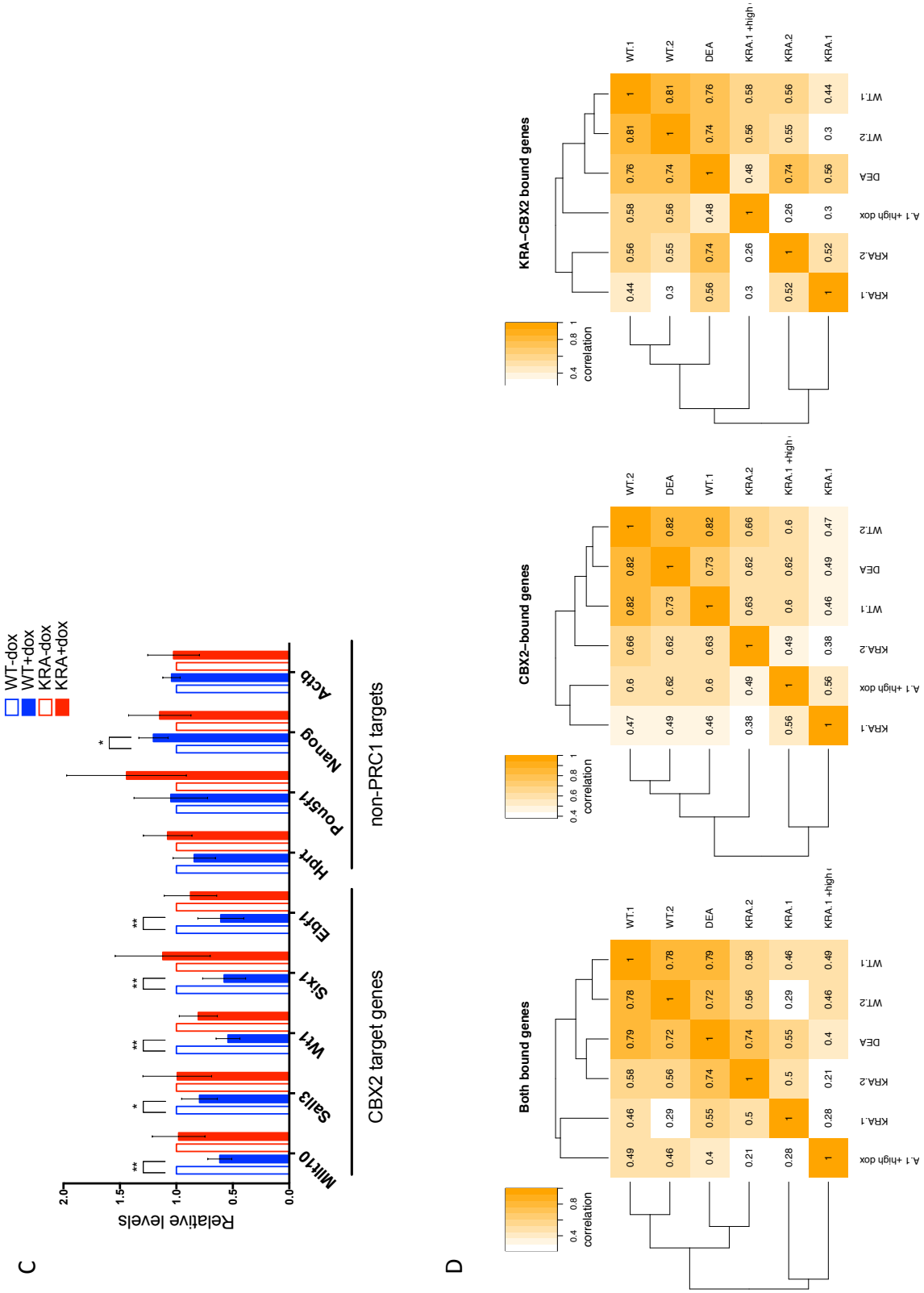


Figure 2.4 (Continued)



the CBX2 mutants to WT, we observed that the target gene set as a whole was affected less by KRA-CBX2 (Figure 2.4A, left panel, right three columns). In contrast, the expression of DEA-CBX2 had effected repression similar to WT-CBX2 (Figure 2.4A, left panel, third column from left). Again, these effects on target gene expression are evident when the RNA seq tracks of individual affected genes were examined (Figure 2.4B). The effects on gene repression were also verified by reverse transcription-qPCR on the cell lines before and after doxycycline treatment (Figure 2.4C). Therefore, we conclude that WT-CBX2 is required for the repression of a subset of genes that it binds. KRA-CBX2 has less repressive activity on these genes compared to WT-CBX2, but DEA-CBX2 has similar activity to WT-CBX2.

We performed hierarchical clustering on the gene expression changes in the cell lines to determine their degree of similarity. The WT- and DEA-CBX2 expressing lines clustered together, while the KRA-CBX2 lines clustered together, away from the WT and DEA-CBX2 lines (Figure 2.4A, left panel).

We repeated the analysis on the genes bound by WT-CBX2 and the genes bound by KRA-CBX2. We saw the same results with both lists, in that there was less repression by KRA-CBX2 compared to WT- and DEA-CBX2, and the KRA-lines clustered separately from the WT- and DEA-CBX2 lines (Figure 2.4A, middle and right panels). Finally, we measured the degree of similarity among the cell lines using Pearson correlation test. We found that the WT- and DEA-CBX2 lines show higher correlation with each other than with the KRA-CBX2 lines (Figure 2.4D), which is consistent with the hierarchical clustering analyses.

We conclude that gene repressive activity is decreased in KRA-CBX2, but is unaffected in DEA-CBX2. Therefore, mutations that affect the *in vitro* activities of CBX2 also affect its function *in vivo.*

PRC1 subunits and histone modifications on chromatin are unaffected by KRA-CBX2

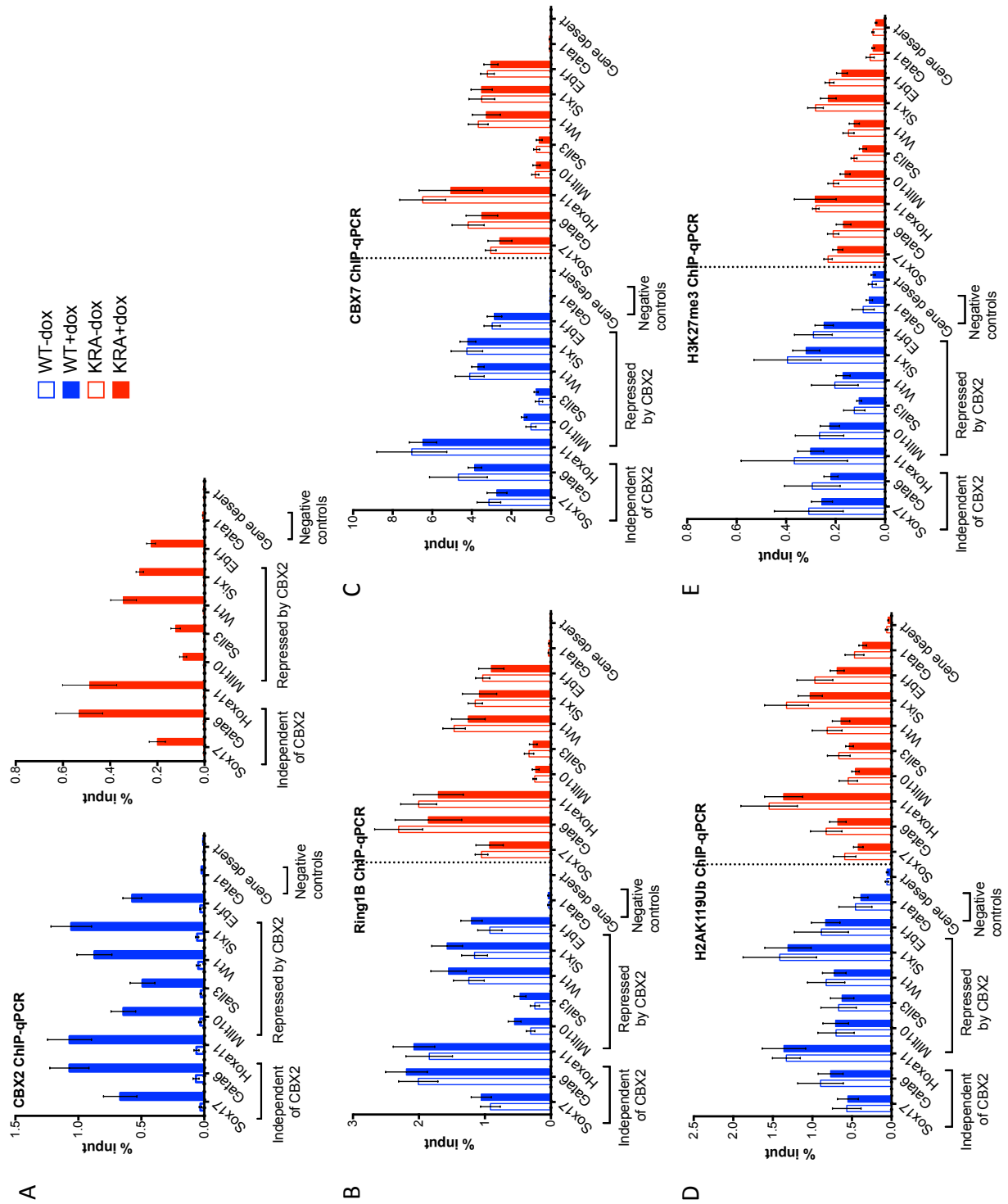
Although the majority of the peaks from the KRA-CBX2 ChIP-seq data overlapped with those from the WT-CBX2 data set, we wanted to know if the differential impact of WT- and KRA-CBX2 on gene expression might be caused by differences in binding specifically at the affected genes. In addition, we also wanted to know whether the other PRC1 components and the histone modifications associated with PcG could have been affected by KRA-CBX2. Thus, we performed ChIP experiments for CBX2, CBX7, Ring1B, H2AUb and H3K27me3 and assayed for enrichment of CBX2-dependent and CBX2-independent regions by qPCR.

From our CBX2-ChIP-qPCR experiments, we observed that KRA-CBX2 is able to bind to both genes that are dependent and independent of WT-CBX2 for repression, similar to WT-CBX2 (Figure 2.5A). Therefore, the lack of repression by KRA-CBX2 is not due to its absence at the affected genes. Notably, the ChIP enrichment for KRA-CBX2 is about half of that for WT-CBX2 which is consistent with our ChIP-seq analysis described earlier (Figure 2.2E).

The lack of repression by KRA-CBX2 could be due to interference with the PRC1 complex on the chromatin. As Ring1B is a central component of all PRC1 family complexes, we took it as a representative of all PRC1 complexes and determined whether its binding on chromatin is affected. We observed that re-introducing the WT- or KRA-CBX2 does not affect Ring1B binding compared to their respective null states (Figure 2.5B). This implies that CBX2-mediated repression is not due to increased recruitment of other PcG proteins to the target genes. It also implies that KRA-CBX2 does not interfere with the binding of these

Figure 2.5 CBX2-dependent repression is not correlated with Ring1B, CBX7, H2AK119Ub and H3K27me3 levels at target genes. ChIP-qPCR analyses for **A.** CBX2, **B.** Ring1B, **C.** CBX7, **D.** H2AK119Ub and **E.** H3K27me3 at genes that are repressed by WT-CBX2, independent of CBX2, and negative control regions. Plotted values are Mean± SD, n≥3. T-tests comparing the means of –dox and +dox at individual genes for Ring1B, CBX7, H2AK119Ub and H3K27me3 data were performed and no statistically significant differences were obtained.

Figure 2.5 (Continued)



proteins. Therefore, we conclude that CBX2, whether WT or KRA, does not affect Ring1B on the chromatin.

CBX7 is the dominant CBX homolog that is expressed in mESCs. We wanted to know whether CBX7 binds differentially at genes that are dependent and independent of CBX2 for repression, which could imply distinct functions for different CBX-containing PRC1 complexes. By CBX7-ChIPqPCR, we observed that CBX7 is present at both CBX2- dependent and independent genes, and that its binding is independent of CBX2 (Figure 2.5C, compare the blue open and solid bars). CBX7 alone is thus insufficient for the full repression of CBX2-dependent genes. This implies that CBX7 and CBX2 are not fully redundant. We also observed that KRA-CBX2 does not affect CBX7 binding (Figure 2.5C, compare the red open and solid bars). Hence, we conclude that CBX7 or CBX7-containing PRC1 is insufficient for full repression of genes that require CBX2 for repression.

Finally, CBX2-dependent repression might be associated with H2AUb andH3K27me3, the two histone modifications that coincide with PcG repressed regions in the genome. We found that re-introducing WT-CBX2 does not increase the levels of H2AUb (Figure 2.5D, compare blue open and solid bars). This implies that CBX2-mediated repression is independent of H2AUb; it also implies that H2AUb without CBX2 is insufficient for the full repression of some genes. We also noted that KRA-CBX2 does not interfere with H2AUb levels, indicating that the mutant CBX2 does not disrupt Ring1B catalytic activity (Figure 2.5D, compare red open and solid bars). Similarly, the re-introduction of WT-CBX2 does not increase H3K27me3 levels (Figure 2.5E, compare blue open and solid bars). This suggests that CBX2-dependent genes are not fully silenced without CBX2 even when H3K27me3 is present at wild type levels. Again, KRA-CBX2 does not interfere with

H3K27me3 levels. Based on these observations, we conclude that repression by CBX2 is independent of Polycomb-associated histone modifications.

Taken together, CBX2 seems to be the direct effector of silencing at a subset of genes. Mutations in the CBX2 compaction domain inhibit its gene silencing activity.

DISCUSSION

In summary, we observed that behaviors of WT-, KRA- and DEA-CBX2 are different in a cellular context, and appeared to be associated with their *in vitro* activities. WT- and DEA-, both of which are active *in vitro,* are able to repress PRC1 target genes, and their protein levels are tightly regulated in the cells. In contrast, KRA-CBX2, which is inactive *in vitro,* is unable to repress CBX2 target genes, and can be tolerated at much higher levels; this corroborates that it is biologically inactive. Our results indicated that the differences between WT- and KRA-CBX2 are not due to the loss of complex formation, chromatin binding or mis-targeting with KRA-CBX2. There are no significant changes in the binding of other PRC1 components or Polycomb-associated histone modifications on the target genes whether in the presence of WT- or KRA-CBX2. This suggests that the lack of gene repression by KRA-CBX2 is due to the absence of an activity that directly affects the target genes. The dominant CBX homolog in mESCs, CBX7, is inactive *in vitro* and does not fully repress PRC1 targets that are CBX2-dependent. Hence, our observations support the hypothesis that chromatin compaction, as exerted by CBX2, is a mechanism by which PRC1 achieves gene repression.

We observed in our study that the number of genes that became repressed in *Cbx2*$^{-/-}$ mESCs following the re-introduction of WT-CBX2 is small, and there were no change in the

expression of the *Hox* genes. Moreover, the magnitude of down-regulation is also modest with most target genes showing less than two-fold change. This observation is consistent with a previous finding that CBX2 is not essential for the role of PRC1 in the mESC state[25]. This led us to ask whether the difference in the WT- and KRA-CBX2 repressive activities has significance in biological functions; it is known that biological systems can tolerate a small range of gene expression levels without any impact.

Cbx2 expression is low in mESCs, and increases as the cells undergo differentiation (Ref. [25]; also our data not shown). This suggests that we may see a larger difference between WT- and KRA-CBX2 activity in a differentiated cell type. We differentiated our mESC lines using several differentiation protocols, and found that the differentiated cells often fail to respond to doxycycline induction for CBX2 expression. The phenomenon is unlikely to be specific to WT- and KRA-CBX2, as others and we had experienced a similar challenge with unrelated doxycycline –inducible cell lines. Therefore, we subsequently developed a new system to further address the impact of the mutations in the compaction domain of CBX2, as described below.

CONCLUSIONS AND NEXT STEPS

Through comparing mESCs that express WT- or KRA-CBX2, we were able to conclude that mutations in the compaction domain disrupt gene repression by CBX2. The magnitude of the effects, however, is small, which prompted us to compare WT- and KRA-CBX2 in a different system.

Two independently generated *Cbx2*$^{-/-}$ mice had been studied and have a range of phenotypes, including anterior-to-posterior transformations that defined the Polycomb

group mutants[32,33]. We reasoned that we would generate mutant mice that express KRA-CBX2 and compare phenotypes to the *Cbx2*$^{-/-}$ phenotypes. With the mutant mice we will also be able to study WT- and KRA-CBX2 in the context of the other CBX homologs that are expressed during development. This is interesting because the compaction domain that is described here is unique to CBX2 amongst the other CBX homologs. This work is described in the next research chapter.

MATERIALS AND METHODS

Cell culture

Cbx2$^{-/-}$ and its derivative doxycycline-inducible mESC lines were cultured on DR4 feeders in mESC media, containing DMEM (Gibco 11995), supplemented with 15% FBS (Tet-System Approved FBS, ES Cell Qualified; ClonTech Laboratories 631157), 1000U/mL LIF (Millipore), non-essential amino acids, GlutaMax , penicillin/streptomycin , 2-β-mercaptoethanol . V6.5, CJ7 and J1 mESCs were cultured on feeders (PMEF-N, Fischer Scientific) in mESC media prepared as above, except that the FBS is ES-certified serum from Hyclone (Thermofischer Scientific 16141079). E14 were culture on 0.2% gelatin-coated plate in mESC media supplemented with Hyclone serum.

Generation of doxycycline-inducible mESC lines

The cell lines were generated using Tet-On® 3G Inducible Expression System (EF1alpha Version) and Tet-On® 3G Vector Set (Bicistronic Version) (Clontech 631167, 631161) according to the manufacturer's protocol. Briefly, the *Cbx2*$^{-/-}$ cells were transfected with the pEF1alpha-Tet3G plasmid and puromycin linear selection marker and selected for puromycin resistance. To test their responsiveness to doxycycline, individual clones were transfected with pTRE3G-Luc and assayed for luciferase activity. A responsive clone was chosen and transfected with the pTRE3G plasmid containing a Kozak sequence followed by the coding sequence for HA-tagged WT, KRA or DEA-CBX2 in MCS1, and the coding sequence for GFP in MCS2, along with linear hygromycin selection marker. Hygromycin-resistant clones were treated with doxycycline and screened for GFP expression by imaging and for expression of CBX2 by Western blot.

Western blot

Whole cell lysates were prepared by lysing cells in RIPA buffer (50mM Tris-HCl pH 8.0, 0.1% SDS, 0.5% sodium deoxycholate, 1% NP40, 150mM NaCl), separated on 4-12% bis-Tris gels, transferred onto nitrocellulose membranes and blotted for PRC1 components using the antibodies listed in the Appendix, Table S1. TBP or H3 protein levels were used as loading controls. Relative protein levels were quantified using ImageJ.

Chromatin Immunoprecipitation (ChIP)

ChIP experiments were performed using the antibodies listed in the Appendix using standard protocols. 3x10^6 cells were used for each ChIP experiment. ChIP enrichments were assessed via quantitative PCR analysis (ChIP-qPCR) or high throughput sequencing (ChIP-seq).

For ChIP-qPCR, the sequences of the primers are listed in the Appendix, Table S3.

For ChIP-seq, the enriched DNA and input DNA were made into sequencing libraries as previously described[34], and sequenced on Illumina HiSeq 2000 instrument, resulting in approximately 30 million of 50bp reads per sample. Reads were aligned against the mm9 reference genome using BWA[35]. Alignments were filtered for uniquely mapped reads and alignment duplicates were removed. Input-normalized coverage tracks were generated using SPP[36].To determine regions of ChIP-Seq tag enrichment, we analyzed tag counts in a 1kb window over the chromosome length with the step of 200bp. Statistical significance of enrichment of ChIP vs. input was estimated using negative binomial distribution, with the estimate of the mean based on the tag counts in input, and the size parameter (s) selected based on manual inspection of resulting peak calls. Adjacent enriched windows separated

by 1kb or less were merged to generate a list of regions with significant enrichment.

Co-immunoprecipitation

Nuclear lysates and immunoprecipitation were prepared and carried out as described in Hashimoto et al with the following modifications: Lysates from 8x10^6 cells were used per IP without pre-clearing, Protein A Dynabeads (Life Technologies, 100-02D) were used to capture the antibody-antigen complex, and the proteins were eluted by heating at 70°C in LDS for 10 minutes. The antibodies used are listed in the Appendix.

Gene expression analyses

RNA-seq: Total RNA was extracted from cells with Trizol, DNaseI-treated and purified using Direct-zol™ RNA MiniPrep (Zymo). mRNA was purified from the RNA and made into cDNA libraries using TruSeq total RNA kit (Illumina). Sequencing libraries were prepared and sequenced as above. Two or three biological replicates were used per sample. The reads were aligned to the mouse transcriptome (mm9 RefSeq annotation) using TopHat[37]. Gene expression counts were calculated using HTSeq v.0.6.0[38]. Calculation of expression values and differential expression analysis were performed using edgeR[39], with a filter for genes with average read counts of logCPM >1from all biological replicates. Library size-normalized bigwig files for IGV genome browser tracks were generated using SPP.

To generate a list of candidate CBX2 target genes, we selected genes with CBX2 binding based on our CBX2 ChIP-seq data from WT- and KRA-CBX2 expressing mESCs and have log fold change values (logFC) with adjusted P-value <0.1, regardless of the magnitude of fold change. These genes and their logFC were used to generate the heat maps with hierarchical clustering using the heat map.2( ) function in R. The Pearson correlations

scores were based on logFC values and calculated using the cor( ) function in R.

Reverse-transcription quantitative PCR (RT-qPCR): Total RNA was prepared as above and reverse transcribed using SuperScript® VILO™ cDNA Synthesis Kit (Life Technologies, 11754-050). qPCR was performed using the Taqman probes indicated in the Appendix, Table S2. B2M levels were used as the internal control for cDNA content. The expression level of a gene in each +dox sample is presented as a relative value to its expression level in the corresponding no dox sample.

REFERENCES

1. Paro, R. & Hogness, D. S. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. Proc. Natl. Acad. Sci. U.S.A. 88, 263–267 (1991).

2. Kennison, J. A. & Tamkun, J. W. Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. Proc. Natl. Acad. Sci. U.S.A. 85, 8136–8140 (1988).

3. Martin, E. C. & Adler, P. N. The Polycomb group gene Posterior Sex Combs encodes a chromosomal protein. Development 117, 641–655 (1993).

4. Rastelli, L., Chan, C. S. & Pirrotta, V. Related chromosome binding sites for zeste, suppressors of zeste and Polycomb group proteins in Drosophila and their dependence on Enhancer of zeste function. EMBO J 12, 1513–1522 (1993).

5. Aranda, S., Mas, G. & Di Croce, L. Regulation of gene transcription by Polycomb proteins. Sci Adv 1, e1500737 (2015).

6. Wang, H. et al. Role of histone H2A ubiquitination in Polycomb silencing. Nat Cell Biol 431, 2799–2810 (2004).

7. Elderkin, S. et al. A Phosphorylated Form of Mel-18 Targets the Ring1B Histone H2A Ubiquitin Ligase to Chromatin. Molecular Cell 28, 107–120 (2007).

8. Stock, J. K. et al. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat Cell Biol 9, 1428–1435 (2007).

9. Pengelly, A. R., Kalb, R., Finkl, K. & Müller, J. Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. Genes & Development (2015). doi:10.1101/gad.265439.115

10. Illingworth, R. S. et al. The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. Genes & Development 29, 1897–1902 (2015).

11. Isono, K. et al. SAM Domain Polymerization Links Subnuclear Clustering of PRC1 to Gene Silencing. Developmental Cell 26, 565–577 (2013).

12. Wani, A. H. et al. Chromatin topology is coupled to Polycomb group protein subnuclear organization. Nat Comms 7, 10291 (2016).

13. Pengelly, A. R., Copur, Ö., Jäckle, H., Herzig, A. & Müller, J. A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. Science 339, 698–699 (2013).

14. Bernstein, E. et al. Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin. Mol. Cell.

Biol. 26, 2560–2569 (2006).

15. Hatano, A., Matsumoto, M., Higashinakagawa, T. & Nakayama, K. I. Phosphorylation of the chromodomain changes the binding specificity of Cbx2 for methylated histone H3. Biochem. Biophys. Res. Commun. 397, 93–99 (2010).

16. Kaustov, L. et al. Recognition and Specificity Determinants of the Human Cbx Chromodomains. Journal of Biological Chemistry 286, 521–529 (2010).

17. King, I. F. G., Francis, N. J. & Kingston, R. E. Native and Recombinant Polycomb Group Complexes Establish a Selective Block to Template Accessibility To Repress Transcription In Vitro. Molecular and Cellular Biology 22, 7919-7928 (2002).

18. Francis, N. J. Chromatin Compaction by a Polycomb Group Protein Complex. Science 306, 1574–1577 (2004).

19. Schloßherr, J., Eggert, H., Paro, R., Cremer, S. & Jack, R. S. Gene inactivation in Drosophiia mediated by the Polycomb gene product or by position-effect variegation does not involve major changes in the accessibility of the chromatin fibre. Molec. Gen. Genet. 243, 453–462 (1994).

20. Boivin, A. & Dura, J.-M. In vivo chromatin accessibility correlates with gene silencing in Drosophila. Genetics (1998).

21. Bell, O. et al. Accessibility of the Drosophila genome discriminates PcG repression, H4K16 acetylation and replication timing. Nat Struct Mol Biol 17, 894–900 (2010).

22. King, I. F. G. et al. Analysis of a Polycomb Group Protein Defines Regions That Link Repressive Activity on Nucleosomal Templates to In Vivo Function. Mol. Cell. Biol. 25, 6578–6591 (2005).

23. Grau, D. J. et al. Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. Genes & Development 25, 2210–2221 (2011).

24. Beh, L. Y., Colwell, L. J. & Francis, N. J. A core subunit of Polycomb repressive complex 1 is broadly conserved in function but not primary sequence. PNAS E1063-1071 (2012).

25. Morey, L. et al. Nonoverlapping Functions of the Polycomb Group Cbx Family of Proteins in Embryonic Stem Cells. Cell Stem Cell 10, 47–62 (2012).

26. O'Loghlen, A. et al. MicroRNA Regulation of Cbx7 Mediates a Switch of Polycomb Orthologs during ESC Differentiation. Cell Stem Cell 10, 33–46 (2012).

27. Boyer, L. A. et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nat Cell Biol 441, 349–353 (2006).

28. Vincenz, C. & Kerppola, T. K. Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. PNAS

105: 16572-16577(2008).

29. Sharp, E. J., Martin, E. C. & Adler, P. N. Directed overexpression of suppressor 2 of zeste and Posterior Sex Combs results in bristle abnormalities in Drosophila melanogaster. Developmental biology 161, 379–392 (1994).

30. Bel, S. et al. Genetic interactions and dosage effects of Polycomb group genes in mice. Development 125, 3543–3551 (1998).

31. Alkema, M. J., van der Lugt, N. M. T., Bobeldijk, R. C., Berns, A. & Van Lohuizen, M. Transformation of axial skeleton due to overexpression of bmi-1 in transgenic mice. Nature 374, 724 (1995).

32. Katoh-Fukui, Y. et al. Male-to-female sex reversal in M33 mutant mice. Nature 393, 688–692 (1998).

33. Coré, N. et al. Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice. Development 124, 721–729 (1997).

34. Bowman et al Multiplexed Illumina sequencing libraries from picogram quantities of DNA. BMC Genomics 14, 466-472 (2013).

35. Li, H., & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14), 1754–1760 (2009).

36. Kharchenko, P. V., Tolstorukov, M. Y., & Park, P. J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nature Biotechnology, 26(12), 1351–1359. (2008).

37. Trapnell, C., Pachter, L., & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics, 25(9), 1105–1111 (2009).

38. Anders S., Pyl P.T., & Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31 (2): 166-169 (2015).

39. Robinson, M. D., McCarthy, D. J., & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26(1), 139–140 (2009).

CHAPTER 3

Mutations in the compaction domain of CBX2 lead to posterior transformations

Contributions:

For the generation of the mutant mice, Gerald Marsischky (Genome Engineering Production Group, Department of Genetics) assembled the ES targeting construct and sgRNAs; the Genome Modification Facility at Harvard University performed the blastocyst and zygotic injections. Sharon Marr (Kingston lab) developed the melt curve analysis in Figure 3.3E and genotyped the mice. Daniel Grau (Kingston lab) performed the *in vitro* nucleosome remodeling and inhibition assays and generated Figure 3.4C. Matthew Schwart (Tabin lab) performed the skeletal analyses and generated the images in Figure 3.5.

ABSTRACT

Mutations in the *Polycomb* group genes lead to anterior-to-posterior transformation phenotypes in flies. These phenotypes defined the *Polycomb* group and are a result of loss of *Hox* gene repression by the Polycomb group proteins. In mammals, there are five homologs to the fly *Pc* gene: *Cbx2, Cbx4, Cbx6, Cbx7,* and *Cbx8*. Yet, only the *Cbx2*$^{-/-}$ mice exhibit the anterior-to-posterior transformation phenotype. Among the CBX proteins, CBX2 has a unique region that is required for polynucleosome compaction *in vitro* and gene repression in cells. We hypothesize that this region and its compaction activity are required for PRC1 function during development for proper body axis patterning. To test the hypothesis, we generated mouse mutants with two different knock-in alleles of *Cbx2* with mutations in the compaction domain, and examined their axial skeletons for morphological changes. We observed that these mice exhibit posterior transformation phenotypes akin to the *Cbx2*$^{-/-}$, but not the other developmental abnormalities. Hence, our results support the hypothesis that the compaction domain in CBX2 is specifically required for a PcG function in development..

INTRODUCTION

The Polycomb group proteins are important epigenetic regulators that are required for cell fate stability. They are conserved throughout the animal and plant kingdoms, and function through maintaining the repression of their target genes. The *Polycomb* group mutants (PcG) in flies were initially identified based on a homeotic transformation phenotype, where the heterozygous mutant males exhibited ectopic sex combs on their second and third sets of legs. The homozygous mutants die early during development and the embryos generally have segmentation defects[1,2]. The phenotypes indicated that the anterior-posterior body axis was disrupted during development in the PcG mutants.

The *Hox* genes are important determinants of the anterior-posterior axis in most metazoans. Their expression patterns along the axis determine the ordered identities of the body segments of the organism[3]. Their expression patterns are initiated by early developmental regulators and are maintained through development to adulthood. The homeotic phenotypes in the PcG mutants arise from ectopic expression of the *Hox* genes in the developing embryo[4,5]. The PcG gene products form multi-subunit complexes that bind to chromatin for transcriptional repression.

The mammalian homologs of the PcG genes were identified based on the sequence identities of the proteins they encode. There are multiple mammalian paralogs for each fly Polycomb group gene[6]. The paralogs have distinct and overlapping functions[7,8]. Here, we focus on the mammalian homologs to the fly PC, referred to as the CBX proteins. CBX2 was previously shown to inhibit chromatin remodeling and compact polynucleosomes *in vitro*[9].

There are five PC homologs in the mouse genome: CBX2, CBX4, CBX6, CBX7 and CBX8. They were identified by virtue of their sequence identities to the chromodomain and C-box domain in the fly PC. With the exception of CBX6, the four CBX proteins were shown to bind H3K27me3 and

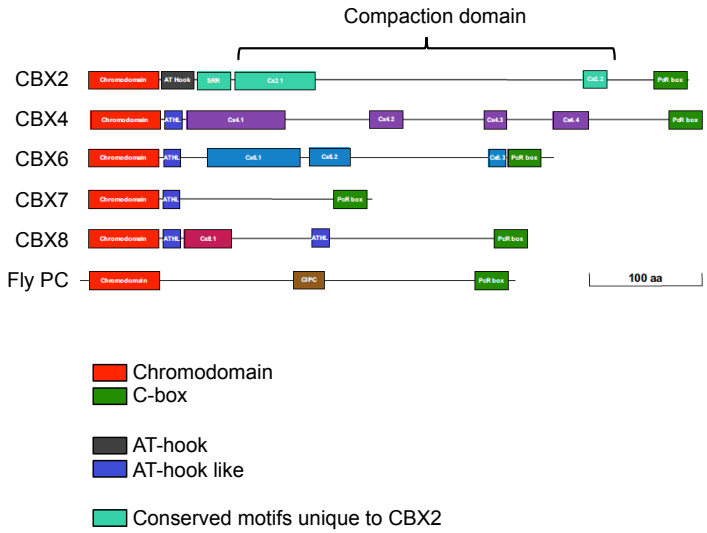
H3K9me3 peptide sequences via their chromodomains[10], interact with Ring1B for PRC1 formation via their C-box domains[11], and localize on chromatin at known PRC1 target genes[12-15]. Hence, they are *bona fide* subunits of the mammalian PRC1. Nonetheless, they have common and distinct characteristics as described below. Thus, depending on which CBX protein is incorporated into PRC1, the activity of the complex is likely to vary. It is unclear how this variation governs the roles of PRC1 in mammalian development, and what features might be unique to the mammalian developmental process compared to flies.

Aside from the chromodomain at the N-terminus, an AT-hook or AT-hook like motif next to the chromodomain and the C-box at the C-terminus, there is little conservation in the sequence and length in the middle portion of the CBX proteins[16] (Figure 3.1A). This variation is thought to confer unique functionalities in the different homologs.

Indeed, the various CBX proteins differ in their abilities to repress chromatin remodeling and compact polynucleosomes *in vitro*[9] (Figure 3.1B). There appears to be a correlation between the length of the variable portion, the amount of positive charge in the protein, and the *in vitro* activity levels. Specifically, the variable portion of CBX2 is the longest and has the highest positive charge, which correlate with the protein's high *in vitro* activity. In contrast, it is the shortest in CBX7, which also is the least positively charged CBX protein and lacks *in vitro* activity. CBX8 has intermediate amounts of positive charge and a corresponding intermediate level of activity.

In addition, the CBX proteins have different subnuclear localization patterns and mobilities when exogenously expressed in mouse embryonic stem cells (mESCs). The differences depend on the non-homologous regions in the protein[14,17]. During mESC differentiation, the assembly stoichiometry of CBX2 on chromatin increases, but not that of CBX7[18]. While these observations

A



B

	Charge	Compaction activity	Organism
CBX2	+32.5	Yes	Mouse
CBX4	Not reported		
CBX6	+16.1	Yes	Frog
CBX7	-7.7	No	Zebrafish
CBX8	+21.3	Yes	Zebrafish

Figure 3.1 Conserved and variable features of the CBX proteins.
A. Illustration of the protein structures of CBX and PC homologs drawn to scale. The region in CBX2 that is required for its *in vitro* polynucleosome compaction activity is indicated. Image is adapted from Senthilkumar and Mishra (2009). **B.** The charge levels and compaction activities of CBX proteins from various organisms. Information is from Grau *et. al* (2012).

support the notion that PRC1-containing different CBX homologs have different roles, the functional significance of the differences remains to be tested.

The expression of CBX2, CBX4, CBX7 and CBX8 are temporally regulated during mESC and hematopoietic stem cell (HSC) differentiation. In mESCs, the expression of CBX7 is the highest and is necessary for self-renewal. As mESCs undergo differentiation, CBX7 is down-regulated while CBX2, 4 and 8 are up-regulated[15]. Even though CBX2, 4 and 8 are expressed at the same developmental stages, they are required for the specification of different cell fates. For example, CBX2 and CBX4 bind distinct sets of target genes in embryoid bodies, and their depletion in teratomas leads to an overrepresentation of cells from the endodermal and mesodermal lineages, respectively. Interestingly, CBX8 is required for activation of developmental regulators in the early differentiation steps, which runs contrary to the conventional role of PRC1 as transcriptional repressors[15]. Similarly for mouse HSCs, only CBX7 is required for self-renewal, while CBX2, 4 and 8 are required for proper differentiation. CBX2 is specifically necessary for lymphopoiesis[19,20]. Hence, CBX7 appears to be associated with sustained cell proliferation and cell fate plasticity, while CBX2, 4 and 8 are associated with cell fate restriction, at least in mice. There is evidence suggesting that the human CBX2 and CBX4 have functions that are contrary to their mouse counterparts[21-23].

The knockout mice for all the CBX homologs have been generated and examined[24-29]. In contrast to flies, in which homozygous *Pc* null is embryonic lethal, the *Cbx* knockout mice are able to complete embryogenesis, suggesting that there is at least some redundancy in function among the CBX proteins during development.

Interestingly, only the *Cbx2*$^{-/-}$ mice exhibit the classic homeotic transformation phenotype that characterized the initial fly *PcG* mutants[24,25]. This suggests that CBX2 has a unique activity that is required for maintaining *Hox* gene repression that cannot be compensated for by the other CBX

homologs. It is unknown whether the lack of compensation is due to the absence of a specific activity in the other homologs or failure to invoke their expression at specific stages of embryogenesis. The *Cbx2*$^{-/-}$ mice also exhibit other phenotypes, including male-to-female sex reversal[25]. This phenotype was reported for a human patient who had two somatic point mutations in *Cbx2*. The mutations are not within the compaction domain defined in our *in vitro* assays, nor did the patient display skeletal abnormalities[30]. This suggests that the protein regions involved in *Hox* gene regulation and sex-determination are distinct in CBX2. In contrast to *Cbx2*$^{-/-}$ mice, the *Cbx4*$^{-/-}$ mice have hematopoietic system defects[31], while the *Cbx7*$^{-/-}$ and *Cbx8*$^{-/-}$ mice do not have obvious developmental abnormalities[26,32].

We reported in Chapter 2 that the residues that are important for the *in vitro* activities of CBX2 are also required for repression of PRC1 target genes in mESCs. Following this, we wanted to assess if the same residues are similarly required for PRC1 function during embryogenesis, specifically, if they can account for the homeotic transformation phenotype that is uniquely exhibited by the *Cbx2*$^{-/-}$ mice. Therefore, we generated two mutant mouse lines with knock-in *Cbx2* alleles that express inactive CBX2 variants. The mutants differ in the number of K-to-A and R-to-A mutations that were introduced. One mutant allele expresses the KRA-CBX2 protein studied in Chapter 2, which contains 23 amino acid substitutions, while the other expresses CBX2 with 13 amino acid substitutions. We examined these mice in detail, with a focus on the homeotic transformation phenotypes.

RESULTS

Generation of *Cbx2*^{23KRA} mutant mice

To determine whether K-to-A and R-to-A mutations in the compaction domain would negatively impact PRC1 function during embryogenesis, we generated a targeting construct to knock-in the mutations into the endogenous *Cbx2* locus (Figure 3.2A). The predicted gene product from this allele is the same KRA-CBX2 variant that was investigated in Chapter 2. It will be referred to as CBX2^{23KRA} from here onwards due to the 23 amino acid substitutions introduced.

The construct was electroporated into V6.5 mESCs, which is a C57BL/6 x 129S4/SvJae F1 hybrid line; the genetic background of the resulting mice will be a point of consideration in the phenotypic analyses later. We screened for homologous recombination events at the endogenous locus by PCR with the primer pairs indicated in Figure 3.2A. The targeted and the wild type alleles give PCR products of different sizes, thereby allowing us to identify mESC clones with successful targeting. The targeted allele is denoted *Cbx2*^{23KRAneo} as the neomycin selection cassette (neo^{R}) is present in its intron 4. An example of the PCR products after gel electrophoresis is shown. We recovered one targeted mESC clone.

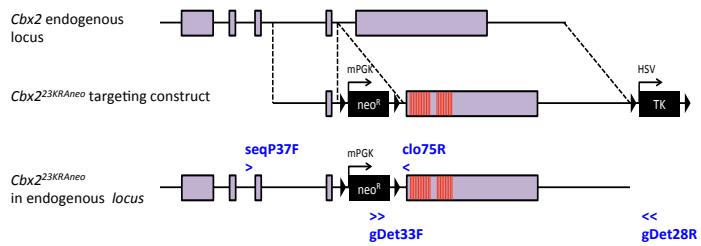
The steps to obtain *Cbx2*^{23KRA} mutant mice are summarized in Figure 3.2B. Briefly, the targeted mESC clone was injected into host blastocysts to generate chimeric mice, which were then crossed with C57BL/6 mice to test for germ line transmission of *Cbx2*^{23KRAneo}. We recovered four chimeric founders with germ line transmission.

To remove neo^{R} from *Cbx2*^{23KRAneo}, heterozygous F1 males, or homozygous males from F1 intercrosses, were mated with *EIIa-Cre* transgenic females. The activity of the CRE protein in the pre-implantation embryos from these crosses removes the neo^{R} that was flanked by loxP sites.

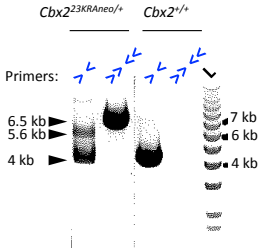
Figure 3.2 Generation of *Cbx2*^{23KRA} mutant mice.
A. Strategy for knocking-in the 23 K-to-A and R-to-A mutations. The endogenous wild type (WT) locus, targeting construct, and targeted locus after homologous recombination are shown. The binding locations of the primers used to screen for successful targeting events are indicated in blue arrowheads. The table presents the expected sizes of the PCR products from the primers. An example gel image of the PCR products after electrophoresis is shown. **B.** Steps in generating *Cbx2*^{23KRA} mutant mice from the targeted mESC clone, which includes mating with Cre-expressing transgenic mice to remove neomycin cassette (neo^{R}) from the targeted *Cbx2* allele. **C.** The *Cbx2*^{23KRA} allele before and after neo^{R} removal, together with the binding locations of the primers used to identify successful CRE-*lox* excision events. The expected sizes of the PCR products and an example gel image after electrophoresis are shown. For sample 1 on this gel image, the presence of a 594bp and a 308bp band indicates that this genomic DNA material (from tail tip) contains *Cbx2*^{23KRA} with and without neo^{R} (denoted as *Cbx2*^{23KRAneo} and *Cbx2*^{23KRA}, respectively). This indicates that Cre-mediated removal of neo^{R} is incomplete in this mouse. For sample 2, only the 594bp band is present. This indicates that the genomic DNA material contains only *Cbx2*^{23KRA}.

Figure 3.2 (Continued)

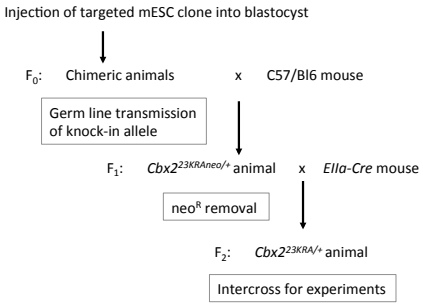
A



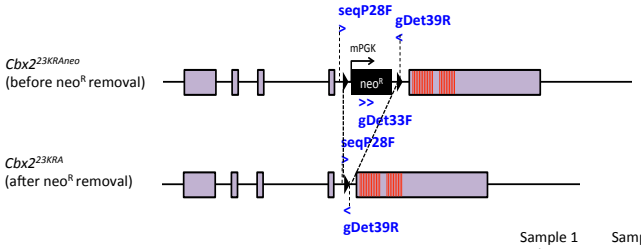
	Primer pairs	seqP37F/ clo75R	gDet33F/ gDet28R
PCR product size	*Cbx2*^{23KRAneo} allele	5.6 kb	6.5 kb
	Cbx2^{+} allele	4 kb	No product



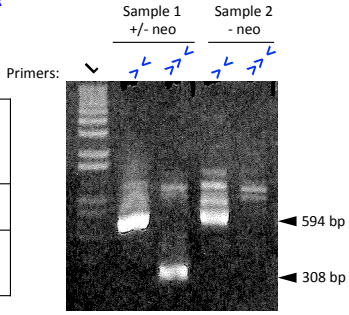
B



C



	Primer pairs	seqP28F/ gDet39R	gDet33F/ gDet39R
PCR product size	*Cbx2*^{23KRAneo} allele	5.6 kb	6.5 kb
	Cbx2^{23KRA} allele	4 kb	No product



Heterozygous F2 progeny from this cross were screened for the removal of neo^R by PCR with the indicated primer pairs (Figure 3.2C). An example of the PCR products after gel electrophoresis is shown. We denote the targeted allele after neo^R removal *Cbx2^{23KRA}*. Only *Cbx2$^{23KRA/+}$* mice were used in subsequent crosses. We noted that neo^R removal could be incomplete in the germ line of these mice and this cannot be detected by genotyping with tail tip genomic DNA. Therefore, we always tested the F3 progeny from these crosses for the presence of neo^R. Only F3 mice that were free of neo^R were retained for further breeding.

The *Cbx2$^{23KRA/+}$* F2 mice were intercrossed, and the F3 progeny were genotyped using a melt curve analysis approach (see Materials and Methods). We were able to obtain live *Cbx2$^{23KRA/+}$*and *Cbx2$^{23KRA/23KRA}$* animals from these crosses.

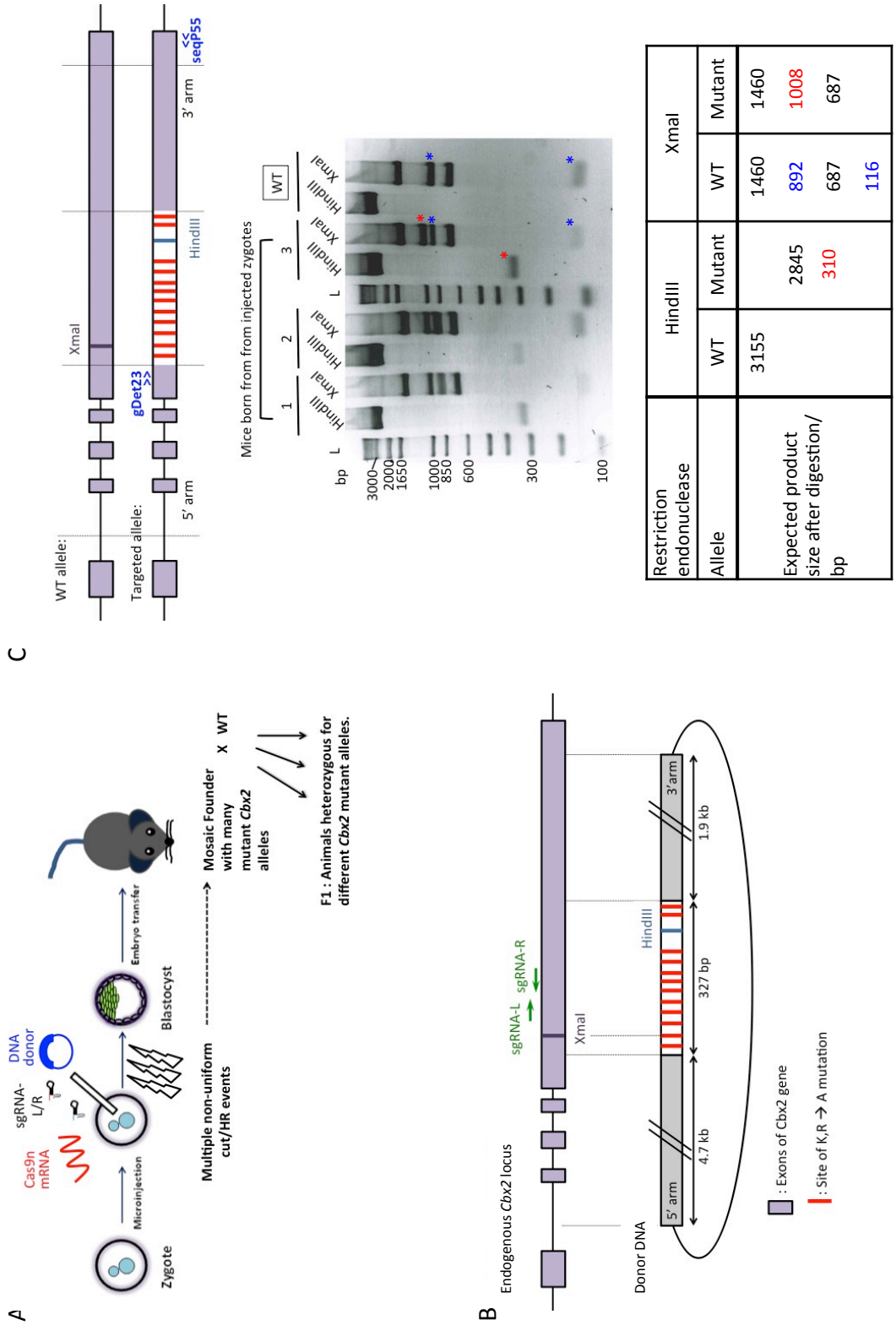
Generation of *Cbx2^{13KRA}* mutant mice

The *in vitro* assays with CBX2 variants indicated that a smaller number of K-to-A and R-to-A substitutions also negatively impacted CBX2 activity levels[9]. We wanted to know if a smaller number of mutations (fewer than 23) would also disrupt PRC1 function during development. We chose to test CBX2^{13KRA} because of its *in vitro* activity levels, which are intermediate between CBX2^{WT} and CBX2^{23KRA}. All the 13 amino acid substitutions in CBX2^{13KRA} are in region 1 of the biochemically defined compaction domain. It appears that this region is critical for CBX2 activity, as a similar number of mutations in region 2 of the compaction domain by themselves had no impact on the activity but enhance the loss of activity in the presence of the mutations in region 1.

To generate a knock-in *Cbx2^{13KRA}* allele, we used a CRIPSR-Cas9-mediated targeting approach as illustrated in Figure 3.3A. A pair of single-guide RNAs (sgRNAs), mRNA encoding the Cas9 nickase, and a plasmid containing the donor sequence for homologous recombination were injected into mouse zygotes (C57BL/6 inbred strain). The schematics of the recognition sites of the

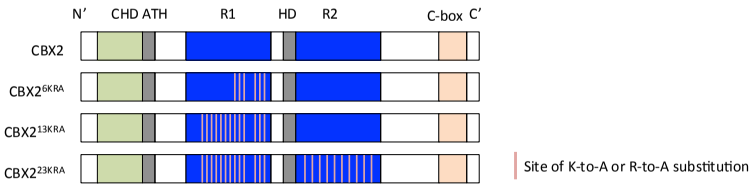
Figure 3.3 Generation of *Cbx2*^{13KRA} mutant mice.
A. The approach for generating *Cbx2*^{13KRA} mutant mice via zygotic injection of the CRISPR-Cas9n system and a DNA donor. Image is adapted from http://jaxmice.jax.org. **B.** Schematics of the endogenous *Cbx2* locus and the composition of the DNA donor used for knocking in the K-to-A and R-to-A mutation sequence. The relative positions on gene of the recognition sites of the sgRNA pair, the sites encoding the mutations, and the restriction sites for RFLP analysis are indicated. **C.** The endogenous WT allele and the expected targeted allele after homologous recombination. The binding locations of the primers used in the PCR-RFLP screening assay are indicated by blue arrowheads. An example of the gel image after RFLP and the fragment sizes that are expected from each allele. Fragments unique to the mutant and WT allele are marked with red and blue asterisks. respectively. In this example, the three mice that arose from injected zygotes had RFLP fragments from both the mutant and WT alleles. This indicates that one of their Cbx2 allele had undergone the anticipated homologous recombination event. **D.** Schematic illustration of the expected gene products from the *Cbx2* mutant alleles in mice generated for this study. The amino acid sequence in their compaction domains are shown. **E.** An example of the melt curve profiles of WT, heterozygous and homozygous mutants (bottom panel, shown for *the Cbx2*^{13KRA} mice). The binding sites of the primers used in this analysis are shown in the top panel. The middle panel are melt curves from control plasmids. The profiles for *Cbx2*^{23KRA} and *Cbx2*^{6KRA} alleles are similar to the profile depicted.

Figure 3.3 (Continued)



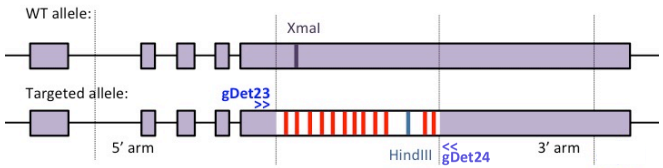
Restriction endonuclease	HindIII		XmaI	
Allele	WT	Mutant	WT	Mutant
Expected product size after digestion/ bp	3155	2845 310	1460 892 687 116	1460 1008 687

D




```
Cbx2_WT          DEEEDDSDLDSKRGPRGRETHPVPQKKAQILVAKPELKDPIRKKRGRKPLPPEQKAARRP 180
Cbx2_6KRA        DEEEDDSDLDSKRGPRGRETHPVPQKKAQILVAKPELKDPIRKKRGRKPLPPEQAAARAP 180
Cbx2_13KRA       DEEEDDSDLDSARGPAGRETHPVPQAKAQILVAAPELKDPIAKARGAKPLPPEQAAARAP 180
Cbx2_23KRA       DEEEDDSDLDSARGPAGRETHPVPQAKAQILVAAPELKDPIAKARGAKPLPPEQAAARAP 180
                 *********** *** ********* ******* ******* * ** ******* *** *

Cbx2_WT          VSLAKVLKTTRKDLGTSAAKLPPPLSAPVAGLAALKAHTKEACGGPSTMATPENLASLMK 240
Cbx2_6KRA        VSLAKVLATTRADLGTSAAKLPPPLSAPVAGLAALAAHTKEACGGPSTMATPENLASLMA 240
Cbx2_13KRA       VSLAKVLATTRADLGTSAAKLPPPLSAPVAGLAALAAHTKEACGGPSTMATPENLASLMA 240
Cbx2_23KRA       VSLAKVLATTRADLGTSAAKLPPPLSAPVAGLAALAAHTKEACGGPSTMATPENLASLMA 240
                 ******* *** ********************** ***********************

Cbx2_WT          GMAGSPSRGGIWQSSIVHYMNRMSQSQVQAASRLALKAQATNKCGLGLDLKVRTQKGGEL 300
Cbx2_6KRA        GMAGSPSRGGIWQSSIVHYMNRMSQSQVQAASRLALKAQATNKCGLGLDLKVRTQKGGEL 300
Cbx2_13KRA       GMAGSPSRGGIWQSSIVHYMNRMSQSQVQAASRLALKAQATNKCGLGLDLKVRTQKGGEL 300
Cbx2_23KRA       GMAGSPSRGGIWQSSIVHYMNAMSQSQVQAASRLALAAQATNKCGLGLDLAVRTQAGGEL 300
                 ********************** ************** ************* **** ****

Cbx2_WT          GGSPAGGKVPKAPGGGAAEQQRGNHSGSPGAQLAPTQELSLQVLDLQSVKNGVPGVGLLA 360
Cbx2_6KRA        GGSPAGGKVPKAPGGGAAEQQRGNHSGSPGAQLAPTQELSLQVLDLQSVKNGVPGVGLLA 360
Cbx2_13KRA       GGSPAGGKVPKAPGGGAAEQQRGNHSGSPGAQLAPTQELSLQVLDLQSVKNGVPGVGLLA 360
Cbx2_23KRA       GGSPAGGKVPAAPGGGAAEQQRGNHSGSPGAQLAPTQELSLQVLDLQSVANGVPGVGLLA 360
                 ********** ************************************* **********

Cbx2_WT          RHAPAKAIPATNPATGKGPGSGPTGANMTNAPTDNNKGEKLTCKATALPAPSVKRDTVKS 420
Cbx2_6KRA        RHAPAKAIPATNPATGKGPGSGPTGANMTNAPTDNNKGEKLTCKATALPAPSVKRDTVKS 420
Cbx2_13KRA       RHAPAKAIPATNPATGKGPGSGPTGANMTNAPTDNNKGEKLTCKATALPAPSVKRDTVKS 420
Cbx2_23KRA       RHAPAAAIPATNPATGKGPGSGPTGANMTNAPTDNNAGEKLTCAATALPAPSVKADTVKS 420
                 ***** ***************************** ****** ********** *****
```

E



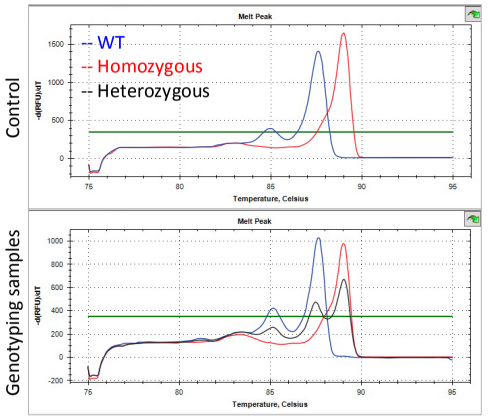


Figure 3.3 (Continued)

sgRNAs and the donor are shown in Figure 3.3B. In addition to the differences in coding sequence, the endogenous *Cbx2* and donor sequences also differ in the presence or absence of the XmaI and HindIII restriction sites, which allows the two sequences to be distinguished by restriction fragment length polymorphism (RFLP).

Mice born from the injected zygotes were screened for homologous recombination events via PCR with the indicated primer pair and RFLP with XmaI and HindIII (Figure 3.3C). Cutting by HindIII and resistance to cutting by XmaI would indicate that homologous recombination has taken place. An example of this analysis is shown in the gel image shown in Figure 3.3C.

Based on the RFLP results, three mice were identified to have homologous recombination events in their genomes. Exon 5 of the *Cbx2* gene from these animals were cloned and sequenced so that they can be characterized in detailed, e.g. to identify small internal insertions or deletions that would not be detected by PCR-RFLP. All three mice were found to have more than two types of *Cbx2* alleles at the endogenous locus, suggesting that they are mosaic. Each type of mutant allele is likely to have arisen from independent CRISPR-mediated cutting events in the developing embryo.

The mosaic founders were crossed to C57BL/6 mice to test for germ line transmission and to segregate the mutant alleles. We recovered several mutant alleles among the progeny, but two are of particular interest: $Cbx2^{13KRA}$, which is encoded by the donor sequence, and a novel $Cbx2^{6KRA}$, that presumably arose from an incomplete homologous recombination event. $Cbx2^{6KRA}$ would give amino acid substitutions in the 8th-13th targeted residues; thus it has half the number of mutations in $Cbx2^{13KRA}$, and a quarter of the mutations in $Cbx2^{23KRA}$ (Figure 3.3D).

$Cbx2^{13KRA/+}$ and $Cbx2^{6KRA/+}$ siblings were intercrossed to produce wild type, heterozygous, and homozygous progeny of the respective mutant alleles for experimentation. The animals were genotyped using the melt curve analysis approach. An example of the melt curve profiles is shown

in Figure 3.3E. We were able to obtain live *Cbx2*$^{13KRA/+}$, *Cbx2*$^{13KRA/13KRA}$ and *Cbx2*$^{6KRA/+}$, *Cbx2*$^{6KRA/6KRA}$ animals.

Expression of the engineered *Cbx2* alleles and the activities of their products *in vitro*

To assess whether the targeted mutant alleles were expressed *in vivo*, we compared the CBX2 protein levels in E10.5 embryos that are homozygous mutant with their wild type littermates by Western blot analysis. We observed that CBX2^{23KRA} and CBX2^{13KRA} protein levels are similar to wild type levels,, while CBX2^{6KRA} is present at 66% of wild type levels (Figures 3.3A, B). We repeated the experiment with additional *Cbx2*$^{6KRA/6KRA}$ samples and obtained similar results (data not shown). The protein levels in *Cbx2*$^{6KRA/+}$ embryos, however, are similar to wild type. This indicates that the *Cbx2*^{23KRA} and *Cbx2*^{13KRA} alleles are expressed normally but *Cbx2*^{6KRA} is expressed at a reduced level.

It was reported that the CBX homologs regulate the expression of one another in mESCs and embryoid bodies[13]. If the CBX2 mutants are compromised in their gene repressive function, the expression of the other CBX homologs may be up-regulated, which could confound the interpretation of the mutant phenotypes. The expression patterns of the other *Cbx* genes were not reported in previously published studies of the *Cbx2*$^{-/-}$ mice[24,25]. Thus, we compared the proteins levels of CBX4, 7 and 8 in the homozygous mutants with wild type embryos. We did not measure CBX6 levels because it is not a PRC1 component[13,14]. We found that CBX8 is present at endogenous levels in all three mutants; CBX7 was not detected in all tested embryos, including the wild type (Figure 3.4A, data for CBX7 is not shown). We could not reliably detect a signal using the anti-CBX4 antibodies available. We had previously performed RNA-seq experiments on fore and hind limb buds of from *Cbx2*$^{13RKA/13KRA}$ and *Cbx2*$^{+/+}$ embryos and observed no difference in their mRNA levels

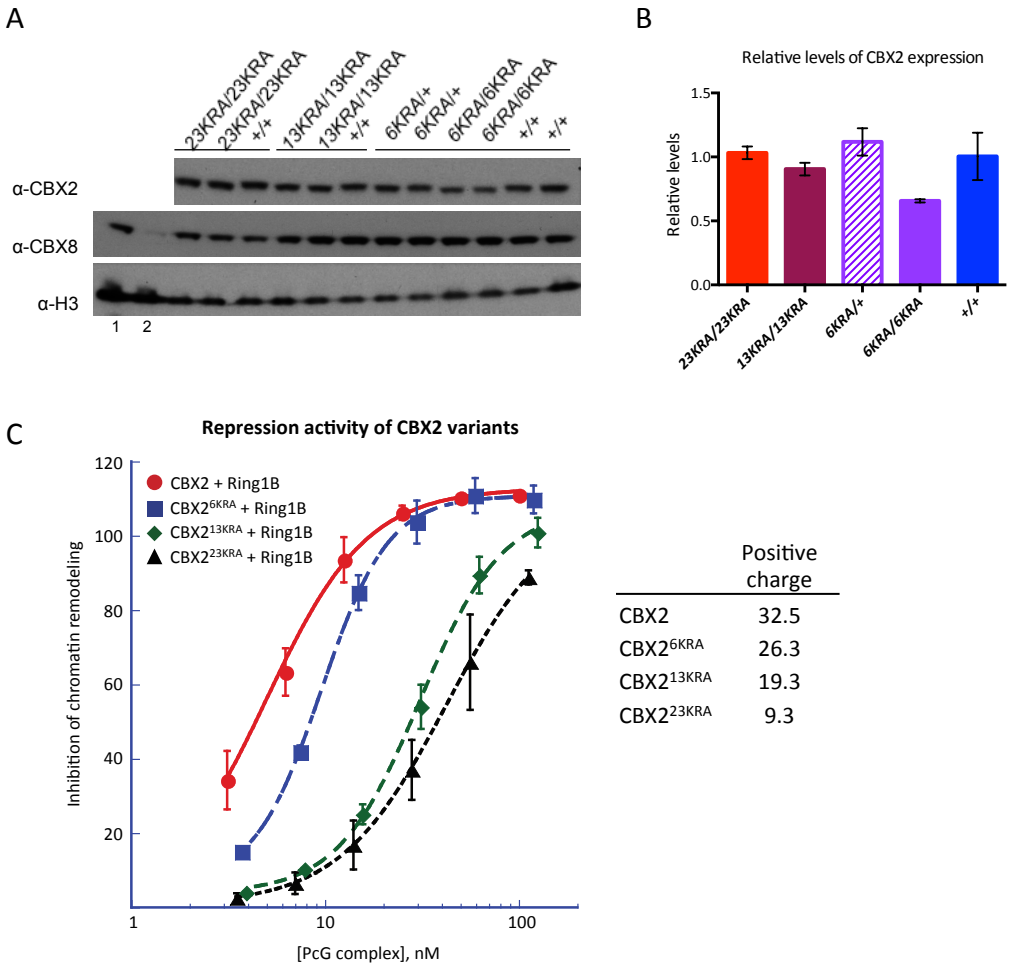


Figure 3.4 The expression of the CBX2^{23KRA}, CBX2^{13KRA} and CBX2^{6KRA} *in vivo* and their *in vitro* activity levels. A. Western blot showing CBX2 and CBX8 in lysates from E10.5 embryos. H3 levels are used as loading controls. Samples with matched genetic backgrounds are marked with horizontal lines. Lanes 1 and 2 are lysates from retinoic-acid differentiated mESCs and un-differentiated mESCs, which serve as positive and negative controls for the anti-CBX8 antibody respectively. **B.** Quantified CBX2 levels in the mutant samples relative to the +/+ samples. The values were obtained by measuring the densities of the WB signals for CBX2 and H3 from Panel A using ImageJ. Plotted values are mean± SD, n=2 for mutants and n=4 for +/+. **C.** The activity level in inhibiting chromatin remodeling for the CBX2 variants in complex with Ring1B. Plotted values are mean ± SEM, n=3. The insert panel shows the positive charge on the CBX2 variants.

for *Cbx4* (data not shown). Therefore, we conclude that the expressions of the other *Cbx* homologs are not affected in the mutants.

CBX2^{6KRA} is a novel CBX2 variant whose *in vitro* activity level was undetermined. In order to be able to associate the phenotypes of the *Cbx2^{6KRA}* mutants with the activity of the protein, we tested CBX2^{6KRA} for its ability to inhibit chromatin remodeling *in vitro.* Based on previous studies we would expect CBX2^{6KRA} to have activity levels that are intermediate between CBX2^{WT} and CBX2^{13KRA}, in accordance to its positive charge. Indeed, this was what we observed (Figure 3.4C). CBX2^{23KRA} is the least active among the protein variants tested, which is consistent with previous findings [9].

Therefore, we have at least two mutant mouse lines, *Cbx2^{23KRA}* and *Cbx2^{13KRA}*, that express CBX2 variants with different extents of K-to-A and R-to-A mutations in the compaction domain. The *Cbx2^{6KRA}* allele has lower CBX2 protein levels. Given that PcG proteins have dose-dependent effects, results from the *Cbx2^{6KRA}* mutants will have to be interpreted with caution. The CBX2 variants from these alleles have *in vitro* activity levels that correspond to the number of mutations introduced. We proceeded to examine the mutant mice for loss-of-function PcG phenotypes to determine the contribution of the compaction domain to PRC1 function in the animals.

Posterior transformation in the skeleton of *Cbx2$^{23KRA/23KRA}$* mice

Posteriorization, or posterior transformation, describes a morphological change where a more anterior body segment assumes the identity of a more posterior segment. Posteriorization is a phenotype for gain-of-function in the *Hox* genes, which in turn can be caused by loss-of-function in the *PcG* genes. In mice, posteriorization is identified by examining the axial skeleton. The mouse vertebrae are divided into five distinct regions, namely the cervical, thoracic, sacral, lumbar and caudal regions. Bones from each segment have unique morphologies and are present in defined

numbers with small variation among different genetic backgrounds. Posteriorization in mice would thus manifest as changes in the vertebrae pattern. Therefore, we stained the skeletons of young adult mice of the *Cbx2*$^{23KR/23KRA}$, *Cbx2*$^{23KR/+}$ and *Cbx2*$^{+/+}$ genotypes for examination.

We found that the vertebrae of the *Cbx2*$^{23KRA/23KRA}$ animals have morphological changes at several positions along the anterior-posterior axis. The changes are all posterior transformations, which are consistent with loss of *Hox* gene repression by PcG proteins during development . These changes included: (1) C7-to-T1: partial transformation of the seventh cervical vertebra to the first thoracic vertebra, as characterized by ectopic ribs at C7, which were often asymmetrical in length and shape. In two of the nine *Cbx2*$^{23KRA/23KRA}$ skeletons examined, the C7 rib on one side is a full rib that fuses to the sternum while the C7 rib on the other side fuses to the rib associated with T1 (Figure 3.5A, B); (2) T1-to-T2: transformation of the first thoracic vertebra to the second thoracic vertebra, as evidenced by the presence of a spinous process on T1 usually restricted to T2 (Figure 3.5A, B); (3) T7-to-T8: unilateral or bilateral transformation of the seventh thoracic vertebra to the eight vertebra, which led to six vertebrosternal ribs instead of the expected seven in wild type mice. In some mutants the seventh rib (R7) misattached to a more posterior position on the sternum (Figure 3.5D); (4) T13-to-L1: the transformation of the thirteenth thoracic vertebra to the first lumbar vertebra, as characterized by the unilateral degeneration of the floating rib normally on T13 (Figure 3.5E, F); (5) L6-to-S1: the transformation of the sixth lumbar vertebra to the first sacral vertebra, which is , identified as the association of the ilium with L6 instead of S1 (Figure 3.5E).

We compared the transformations in the *Cbx2*$^{23KRA/23KRA}$ mice to those previously reported for the *Cbx2*$^{-/-}$*, Phc2*$^{-/-}$ *and Bmi1*$^{-/-}$ *mice* (Table 3.1)[7,24,33,34]. We note that the reported penetrance of the transformations may not be directly comparable among the different mutants as the phenotypes could be scored by different criteria. Furthermore, the mice are in dissimilar genetic backgrounds.

Figure 3.5 Posterior transformations in *Cbx2*^{23KRA} and *Cbx2*^{13KRA} mutant mice.
A. Lateral views of the cervical vertebrae and the first two thoracic vertebrae. The ectopic cervical ribs (asterisks) and spinous processes on T1 (open arrow heads) on the mutants are evidences for posteriorization in the axial skeleton. **B.** Disarticulated C7, T1 and T2 bones from *Cbx2*$^{23KRA/23KRA}$ and matching WT control animals. The transformed bones from the mutants have intermediate morphologies. Mutants exhibit longer cervical ribs (black arrow heads) than the rib anlanges in WT controls. The spinous process on T1 (open arrow head) and a spurious rib extending from the rib on T1 (asterisk) are marked. **C.** Disarticulated C7 bones from *Cbx2*$^{13KRA/13KRA}$ and matching control animals. The cervical ribs in the mutants smaller compared to those in *Cbx2*$^{23KRA/23KRA}$ (Panel B). **D.** Sternal defects. R7 in mutants are often mis-attached or un-attached to the sternum (asterisks). Black arrow heads depict the expected positions of splits in the sternum segment. The absence of the split leads to five sternum segments instead of six. **E.** Transformations in the lumbar and sacral vertebrae. T13-to-L6 transformations are marked with asterisks. **F.** Disarticulated T13 bones from *Cbx2*$^{23KRA/23KRA}$ and matching WT control. The right rib in the mutant sample is degenerated.

Figure 3.5 (Continued)

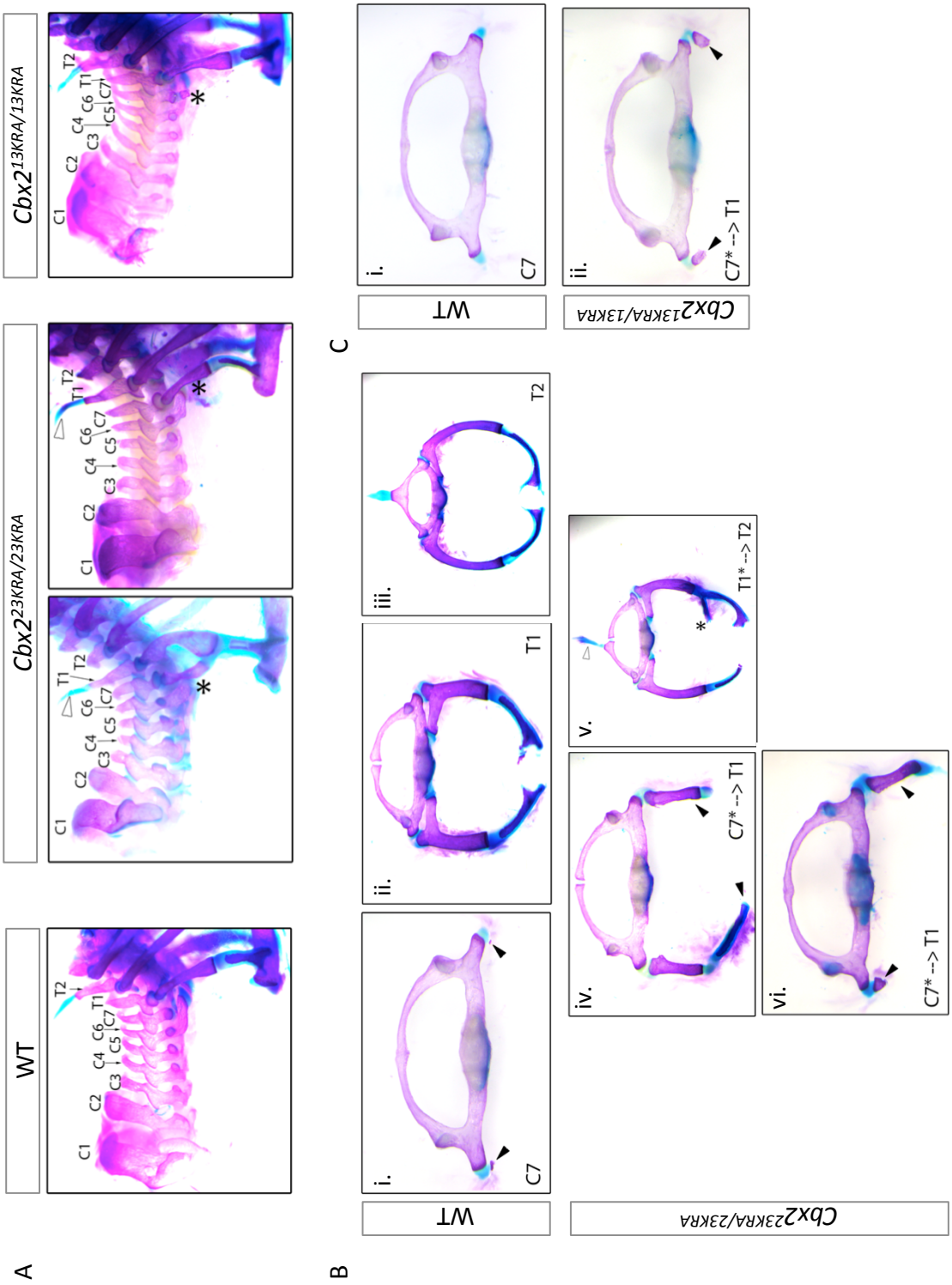


Figure 3.5 (Continued)

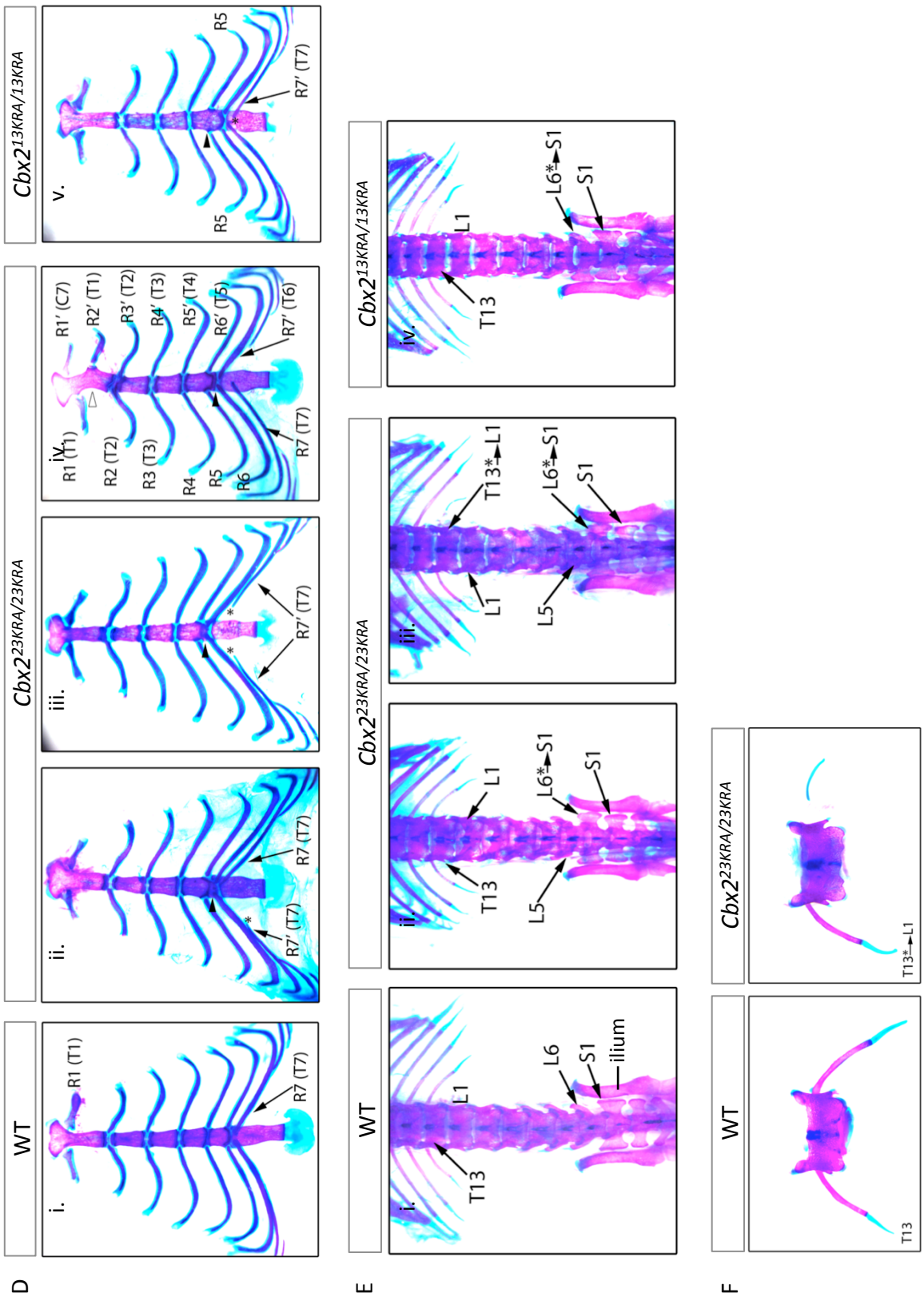


Table 3.1 Skeletal abnormalities in *Cbx2*^{23KRA} and other PRC1 null mutants

	Cbx2 $^{+/+}$ (n=9)	*Cbx2* $^{23KRA/+}$ (n=8)	*Cbx2* $^{23KRA/23KRA}$ (n=9)	*Cbx2* $^{-/-}$ *(Cbx2tm1Ykf)* (n=14)	*Cbx2* $^{-/-}$ *(Cbx2tm1Cim)* (n=13)	*Phc2* $^{-/-}$ (n=6)	*Bmi1* $^{-/-}$ (n=15)
Exocipital-C1 fusion	0	0	0		13 (100%)		
C1-to-C2	0	0	0	14 (100%)	13 (100%)	6 (100%)	
C7-to-T1[1]	0	0	5 (56%)	14 (100%)[2]		4 (67%)	9 (60%)
T1-to-T2	0	1 (13%)	5 (56%)			6 (100%)	
T13-to-L1	0	0	2 (22%)	14 (100%)		5 (83%)	13 (73%)
L6-to-S1 unilateral	1 (11%)	0	1 (11%)		2 (15%)		2 (13%)
bilateral	0	1 (13%)	7 (78%)	14 (100%)[3]		4 (67%)[4]	9 (60%)
Mis-attached/unattached R7 unilateral	0	0	2 (22%)				3 (23%)
bilateral	0	0	4 (44%)		13 (100%)		9 (62%)
5 sternebrae segments	2 (22%)	4 (50%)	8 (89%)				

Table 3.2 Skeletal abnormalities in *Cbx2*^{13KRA} and *Cbx2*^{6KRA} mutants

	Cbx2 $^{+/+}$ (n=2)	*Cbx2* $^{13KRA/+}$ (n=7)	*Cbx2* $^{13KRA/13KRA}$ (n=9)	*Cbx2* $^{+/+}$ (n=7)	*Cbx2* $^{6KRA/+}$ (n=7)	*Cbx2* $^{6KRA/6KRA}$ (n=8)
C7-to-T1[1]	0	3 (43%)	8 (89%)	7 (100%)	6 (86%)	8 (100%)
T1-to-T2	0	0	2 (22%)	1 (14%)	1 (14%)	4 (50%)
T13-to-L1	0	0	0	0	0	0
L6-to-S1 unilateral	0	1 (14%)	3 (33%)	1 (14%)	2 (29%)	2 (25%)
bilateral	0	1 (14%)	2 (22%)	0	2 (29%)	6 (75%)
Mis-attached/unattached R7 unilateral	0	0	3 (33%)	0	0	3 (38%)
bilateral	0	0	0	0	0	0
5 sternebrae segments	1 (50%)	5 (71%)	9 (100%)	3 (43%)	7 (100%)	8 (100%)

[1] There is phenotypic variation in the C7-to-T1 transformation in the *Cbx2*^{23KRA}, *Cbx2*^{13KRA} and *Cbx2*^{63KRA} mutants. Refer to Table 3.3 for a more detailed classification.
[2] This includes C6-to-T1, in addition to C7-to-T1, changes.
[3] This includes L5-to-S1, in addition to L6-toS1, changes.
[4] This was reported without specifying whether the changes are unilateral or bilateral.

The transformations observed in *Cbx2*$^{23KRA/23KRA}$ overlap with those found in *Cbx2*$^{-/-}$ (Table 3.1). This indicates that the *Cbx2*^{23KRA23KRA} resembles *Cbx2*$^{-/-}$. The transformations are partially penetrant in *Cbx2*$^{23KRA/23KRA}$ mice; in contrast, they were reported to be fully penetrant in the *Cbx2*$^{-/-}$ mice. This may suggest that *Cbx2*^{23KRA} is hypomorphic, which is consistent with our molecular data from Chapter 2, where we showed that the K-to-A and R-to-A mutations do not prevent CBX2^{23KRA} from being incorporated into PRC1 and binding chromatin. This could also explain the absence of the exocipital-C1 fusion and C1-to-C2 transformations which were reported for the *Cbx2*$^{-/-}$ mice. Nevertheless, the phenotypic variations could also be due to the differences in genetic backgrounds between our mutant line and the two *Cbx2*$^{-/-}$ mouse lines.

Based on this analysis, we conclude that the mutations in the compaction domain of CBX2 mimic loss-of-function in *Cbx2*. These observations support our hypothesis that activity in the compaction domain is important for CBX2 activity during embryogenesis.

Posterior transformation in the skeleton of *Cbx2*$^{13KRA/13KRA}$ mice

The *in vitro* activity level of CBX2^{13KRA} is between that of CBX2^{WT} and CBX2^{23KRA} (Figure 3.4C). We wanted to know if the *Cbx2*$^{13KRA/13KRA}$ mice have transformation phenotypes, and if so, are they less severe compared to the *Cbx2*$^{23KRA/23KRA}$ mice, as the biochemical data would predict. Thus, we similarly analyzed skeleton preparations of *Cbx2*$^{13KRA/13KRA}$, *Cbx2*$^{13KRA/+}$ and *Cbx2*$^{+/+}$ animals.

We found that the *Cbx2*$^{13KRA/13KRA}$ mice exhibited four out of five of the posterior transformations observed in the *Cbx2*$^{23KRA/23KRA}$ mice (Table 3.2). They did not have additional axial skeletal changes that were not seen in the *Cbx2*^{23KRA}mutants. Thus, the *Cbx2*^{13KRA} and *Cbx2*^{23KRA} alleles lead to similar abnormalities in the axial skeleton.

To determine whether there is a difference in the severity of the phenotypes between the

Cbx2^{13KRA} and *Cbx2*^{23KRA} mutants, we compared the extent of morphological change and the frequency of occurrence of each transformation, taking into consideration that the two lines are in dissimilar genetic backgrounds.

The C7-to-T1 transformations are more dramatic in *Cbx2*$^{23KRA/23KRA}$ than in *Cbx2*$^{13KRA/13KRA}$. For the *Cbx2*$^{23KRA/23KRA}$, the ectopic cervical ribs are longer and in some instances become articulated to the sternum or T1 rib (Figure 3.5B, panels iv and vi). In contrast, the cervical ribs in *Cbx2*$^{13KRA/13KRA}$ resemble nubbins (Figure 3.5C, panel ii). Moreover, two *Cbx2*$^{23KRA/23KRA}$ animals (not the ones with the cervical ribs attached to the sternum) have additional defects in their cervical vertebra, which are C2-C3 and C4-C5 fusions, respectively. Only one *Cbx2*$^{13KRA/23KRA}$ has an additional cervical vertebra defect, which was a missing C3. We noted that the *Cbx2*$^{23KRA/23KRA}$ and *Cbx2*$^{13KRA/13KRA}$ may not be directly comparable for transformations in the cervical region because of differences in genetic background. We observed that the wild type littermates of the *Cbx2*^{23KRA} strain have tiny rib anlages at C7 (Figure 3.5B, panel i). This is presumably due to its hybrid genetic background[35]. In contrast, the wild type littermates of the *Cbx2*$^{13KRA/13KRA}$ animals examined do not have extraneous bones associated with C7 as expected (Figure 3.5C, panel i). Despite the differences between their wild-type littermates, the abnormalities in the cervical vertebrae of *Cbx2*$^{23KRA/23KRA}$ are clearly more severe than those in *Cbx2*$^{13KRA/13KRA}$. Going forward, we will be measuring the length of the C7 ribs in the animals to obtain a quantitative description of the phenotype.

For the T1-to-T2 transformation, the frequency of occurrence is higher in the *Cbx2*$^{23KRA/23KRARA}$ mutants than in the *Cbx2*$^{13KRA/23KRA}$ mutants (56% vs. 22%). Notably this transformation was not found in the wild type animals from either genetic background. It was observed in one *Cbx2*$^{23KRA/+}$ but not in any of the *Cbx2*$^{13KRA/+}$ examined. The T13-to-L13 transformation was found only in *Cbx2*$^{23KR/23KRAA}$ but not in *Cbx2*$^{13KRA/13KRA}$. There is no difference in the wild type and heterozygous mutants from the two strains. Based on these two phenotypes, the

Cbx2^{23KRA} mutation is more severe than the *Cbx2*^{13KRA} mutation.

The L6-to-S1 transformation can occur unilaterally, where the L6 vertebra assumes both lumbar and sacral characteristics (Figure 3.5E, panel iv), or bilaterally, where the L6 appears identical to S1 (Figure 3.5E, panels ii and iii). For *Cbx2*$^{23KRA/23KRA}$, this transformation is predominantly bilateral. For *Cbx2*$^{13KRA/13KRA}$, the frequencies for both types of changes are similar. Overall, taking into consideration both unilateral and bilateral changes, the frequency of the L6-to-S1 phenotype has higher penetrance in *Cbx2*$^{23KRA/23KRA}$ than in *Cbx2*$^{13KRA/13KRA}$ (89% vs. 55%). Similarly, the mis-attachment of R7 can be unilateral (Figure 3.5D, panels ii, iv, v) or bilateral (Figure 3.5D, panel iii). For *Cbx2*$^{13KRA/13KRA}$, the frequency of this change being bilateral is higher than it being unilateral (44% vs. 22%). In contrast, only unilateral changes were found in *Cbx2*$^{13KRA/13KRA.}$. Again, the overall frequency of this phenotype is higher in *Cbx2*$^{23KRA/23KRA}$ than in the *Cbx2*$^{13KRA/13KRA}$ (66% vs. 33%). Thus, the *Cbx2*^{23KRA} mutation is also more severe based on these two phenotypes.

Therefore, we conclude that the skeletal abnormalities collectively indicate that the *Cbx2*^{23KRA} mutation causes more severe posterior transformations than the *Cbx2*^{13KRA} mutation. This correlates with *Cbx2*^{23KRA} being less active than *Cbx2*^{13KRA} in our *in vitro* assays.

We want to highlight that the posterior transformations in the *Cbx2*^{23KRA} and *Cbx2*^{13KRA} mutants are highly similar, albeit with different degrees of severity. This similarity is despite the two strains having different genetic backgrounds and were generated using distinct approaches and targeting constructs. This suggests that an intact compaction domain is essential for PRC1 functions as *Hox* gene regulators during development.

Posterior transformation in the skeleton of *Cbx2*$^{6KRA/6KRA}$ mice

We also analyzed the skeleton preparations of *Cbx2*$^{6KRA/6KRA}$, *Cbx2*$^{6KRA/+}$and *Cbx2*$^{+/+}$ animals.

As CBX2^{6KRA} is more active than CBX2^{13KRA} *in vitro*, we predicted that the *Cbx2*^{6KRA} mutants to have less severe homeotic phenotypes than the *Cbx2*^{13KRA} mutants. However, contrary to our predictions, not only did the *Cbx2*$^{6KRA/6KRA}$ animals have all the homeotic transformations observed in the *Cbx2*^{13KRA13KRA} animals, the frequencies of occurrence were also higher (Table 3.2). A possible explanation for this could be the lower level of CBX2 protein in the *Cbx2*^{6KRA} animals (see Figure 3.4B). However, this does not explain why the transformations are also observed more frequently in the *Cbx2*$^{6KRA/+}$ animals than in the *Cbx2*$^{13KRA/+}$ animals. Curiously, two out of the nine wild type control animals have C7-to-T1 and T1-to-T2 transformations, respectively, that resemble the phenotypes of *Cbx2*$^{23KRA/23KRA}$. In addition, six out of seven of the wild type animals have small cervical ribs, similar to those observed for the wild type littermates for the *Cbx2*^{23KRA} strain (Table 3.3). We expected the wild type controls for the *Cbx2*^{6KRA} and *Cbx2*^{13KRA} strains to be the same because they are in the same genetic background (C57BL/6). This led us to hypothesize that there is background mutation in the *Cbx2*^{6KRA} strain that could be contributing to the phenotype. To test this hypothesis, we are collecting *Cbx2*$^{6KRA/6KRA}$ animals from a different founder animal for analysis. Hence, we are hitherto unable to conclude whether the *Cbx2*^{6KRA} mutation leads to similar or more severe homeotic phenotypes compared to *Cbx2*^{13KRA}.

Survival, growth and fertility of the *Cbx2*^{23KRA}, *Cbx2*^{13KRA} *and Cbx2*^{6KRA} mutants

In addition to sternal defects and posterior transformation in the axial skeleton, which are the classic PcG phenotypes, the *Cbx2*$^{-/-}$ mice have additional phenotypes as described below. We examined our mutant mice for these abnormalities to determine if they are similarly affected as the *Cbx2*$^{-/-}$.

***Post-natal lethality**: Cbx2*$^{-/-}$ exhibit 90% post-natal lethality by 6-weeks of age. In contrast, there is no selective early death among the *Cbx2*^{23KRA}, *Cbx2*^{13KRA} *and Cbx2*^{6KRA} homozygous mutants.

Table 3.3 C7-to-T1 transformations in *Cbx2 23KRA*, *Cbx2 13KRA* and *Cbx2 6KRA* animals

	Strain	***Cbx2 23KRA***			***Cbx2 13KRA***			***Cbx2 6KRA***		
	Genotype	*+/+* (n=9)	*23KRA/+* (n=8)	*23KRA/23KRA* (n=9)	*+/+* (n=2)	*13KRA/+* (n=7)	*13KRA/13KRA* (n=9)	*+/+* (n=7)	*6KRA/+* (n=7)	*6KRA/6KRA* (n=8)
Size of C7 rib	None	0	1 (13%)	0/9	2 (100%)	4 (57%)	1 (11%)	0	1 (14%)	0
	Tiny/Small	9 (100%)[5]	7 (88%)	4 (44%)	0	3 (33%)	8 (89%)	6 (86%)	3 (43%)	6 (75%)
	Medium/Large	0	0/8	5 (56%)	0	0	0	1 (14%)	3 (43%)	2 (25%)

[5] *Cbx2 23KRA/23KRA* animals with C7-rib observation that are tiny/small in Table 3.3 were scored as no phenotype for Table 3.1; only animals that have long cervical ribs as shown in Figure 3.4B (panels iv and vi) were scored as positive for the phenotype in Table 3.1.

The number of progeny from successive heterozygote crosses is shown in Figure 3.6A. The pups were counted when they were born, genotyped between postnatal day 5 and 7 (P5-7) and examined again at weaning (P21). We observed no significant under-representation of the homozygous progeny for all three mutant alleles.

Reduced body weight*: Cbx2*$^{-/-}$ mice exhibit growth retardation - a few days after birth they weigh 75% less compared to wild type littermates. However, these phenotypes were not observed in *Cbx2*^{23KRA}*, Cbx2*^{13KRA} *and Cbx2*^{6KRA} homozygous mice (Figure 3.6B). Therefore, we conclude that the mutations in the compaction domain do not affect growth of the animals.

Sex-reversal*:* The *Cbx2* gene was shown to be important for sex determination in mice and humans (katoh 1998, Lauber 2009). An individual with mutations in the *Cbx2* gene can be genotypically XY but have female external genitalia. In mice, *Cbx2*$^{-/-}$ XY and XX animals are sterile. For all the *Cbx2*^{23KRA}*, Cbx2*^{13KRA} *and Cbx2*^{6KRA} homozygous male and female animals tested, we observed that they have external genitalia that matched their chromosomal sex. Both male and female homozygous mutants are also fertile and able to give viable offspring. Hence, we conclude that the K-to-A and R-to-A mutations do not affect the role of *Cbx2* in sex determination.

Abnormalities in soft tissue organs*:* Finally, we performed histopathology examination on the homozygous mutant mice. Emphasis was given to the colon, spleen, thymus and gonads, given that the *Cbx2*$^{-/-}$ mice develop abnormalities in these organs. We observed no obvious defects in the organs of our mutant animals.

In summary, we observed that the *Cbx2*$^{23KRA/23KRA}$*, Cbx2*$^{13KR/13KRA}$ *and Cbx2*$^{6KRA/6KRA}$animals exhibit a subset of the phenotypes found in the *Cbx2*$^{-/-}$ mice. This indicates that the mutations in the compaction domain disrupted a specific function of the protein. Through examining our mutant mice, we were able to uncouple the roles of CBX2 during development, specifically linking the

A

Outcome	Expected #	Observed #		
+/+	42	38		
23KRA/+	84	90		
23KRA/23KRA	42	40	P value:	0.64
TOTAL	168	168		

Outcome	Expected #	Observed #		
+/+	54.75	59		
13KRA/+	109.5	109		
13KRA/13KRA	54.75	51		
TOTAL	219	219	P value:	0.74

Outcome	Expected #	Observed #		
+/+	43.25	51		
6KRA/+	86.5	82		
6KRA/6KRA	43.25	40		
TOTAL	173	173	P value:	0.39

B

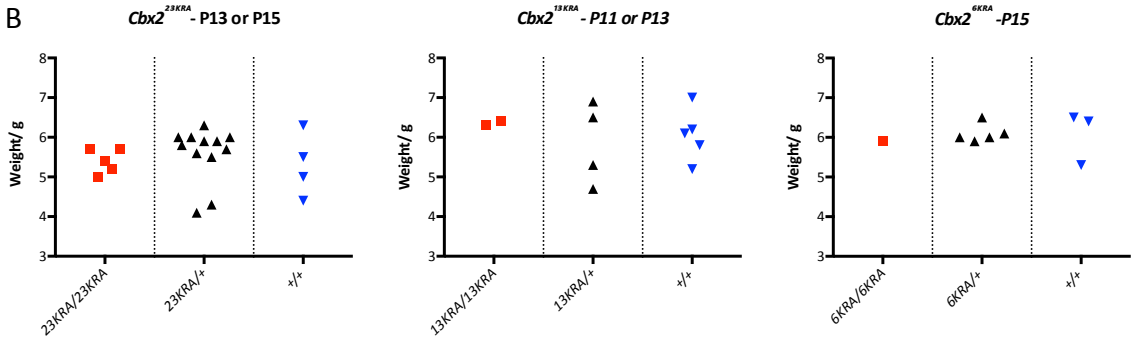


Figure 3.6 No lethality or growth retardation in the *Cbx2*^{23KRA}, *Cbx2*^{13KRA}, *Cbx2*^{6KRA} mutants. **A.** Chi-square test on the number of progeny from heterozygous intercrosses of the mutants. **B.** The weight of the homozygous mutants in comparison to their heterozygous and WT littermates at the indicated ages.

requirement for an intact compaction domain to PcG function in patterning the anterior-posterior axis.

DISCUSSION

In summary, we generated two mutant mice with the knock-in alleles *Cbx2*^{23KRA} *and* *Cbx2*^{13KRA}, respectively, which express CBX2 variants with different number of mutations in the compaction domain. We observed that both *Cbx2*$^{23KRA/23KRA}$ and *Cbx2*$^{13KRA/13KRA}$ animals have posterior transformation phenotypes that are similar to those in *Cbx2*$^{-/-}$ animals, indicating that the mutations lead to loss of function in CBX2. The phenotype appears to be more severe in *Cbx2*$^{23KRA/23KRA}$ than in *Cbx2*$^{13KRA/13KRA}$, and this corresponds to the *in vitro* activity levels of their respective protein products. Our mutants exhibited the PcG phenotype of the *Cbx2*$^{-/-}$ animals, but not the other phenotypes. This indicates that the mutations affected a specific activity in CBX2, which is consistent with our molecular analyses in the previous chapter which indicated that the mutations neither affected CBX2 incorporation into PRC1 nor its localization to chromatin. This alleviated concerns that the introduction a relatively large number of amino acid substitutions (23 and 13, respectively) may affect the protein and the PcG system non-specifically.

Given the extensive knowledge of the field on PcG mutants and homeotic transformation phenotypes, the posterior transformations in our mutants are almost certainly due ectopic *Hox* gene expression. Nonetheless, we have not directly demonstrated this. Thus, a future experiment would be to compare *Hox* gene expression patterns in mutant and wild type embryos by whole mount *in situ* hybridization analyses.

We anticipated that the effects of the mutations on gene expression would be larger in differentiated cell types compared to what was observed in mESCs in the previous chapter. A cellular system in which transcriptional regulation is significantly perturbed by the mutations would be invaluable for future studies that aim to determine the chromatin changes that accompany gene repression by CBX2. To identify such a cellular system, we have performed gene expressions analyses on tissues isolated from mutant and wild type embryos, and primary cell lines that were derived from the embryos. Specifically, we performed RNA-seq experiments on the fore- and hindlimb buds of E10.5 *Cbx2*$^{13KRA/13KRA}$ and wild type embryos. The choice of material was based on the high levels of *Cbx2* expression in these tissues[36]. However, we did not observe obvious differences in the gene expression patterns between the mutant and wild type samples (data not shown). As CBX2 is expressed in primary fibroblasts, we also compared fibroblasts derived from *Cbx2*$^{13KRA/13KRA}$ and wild type embryos for the expression of selected *Hox* genes and *Cdkn2a* (a PRC1 target) by reverse transcription-qPCR. We also did not detect significant differences between the mutant and wild type cells (data not shown). Therefore, we have yet to directly demonstrate that mutating the compaction domain of CBX2 would lead to loss of repression at PRC1 target genes in differentiated cells.

Finally, we acknowledge that our characterization of our mutant mice is far from exhaustive. An interesting phenotype to examine would be the function of the immune system in the mutants. The temporal expression of the CBX homologs is tightly regulated during the development of the hematopoietic system, and CBX2 is specifically required for lymphopoiesis[20,21]. It would be interesting to determine whether the compaction domain, which is unique to CBX2, is important for the development of this particular cell lineage.

CONCLUSIONS AND FUTURE DIRECTIONS

Through the *Cbx2* mutant mice generated, we were able to link the requirement for an intact compaction domain to the genetically defined PcG function during embryogenesis. The same residues in CBX2 are required for chromatin remodeling inhibition and polynucleosome compaction *in* vitro, and for repression of PRC1 targets in mESCs. Taken together, our observations indicate that the *in vitro* activities of CBX2 could be one of the mechanisms by which PRC1 achieve stable gene silencing *in vivo.* This leads to several immediate questions, including the following:

1. Is there a chromatin state in the nucleus that is analogous to the compacted polynucleosomes seen under electron microscopy? How might we assay for this state and how might it prevent transcription?
2. Are the identities of the K and R residues or the positive charge they carry that are the critical features for CBX2 function? In other words, is the nature of the CBX2 and nucleosome contacts driven by specific interactions between chemical groups, or by non-specific electrostatic interactions?
3. How might CBX2 be involved in propagating the repressed state through cellular divisions?

We discuss how we might consider these questions in the next chapter.

MATERIALS AND METHODS

All animal procedures were performed according to NIH guidelines and approved by the Committee on Animal Care at Massachusetts General Hospital and Harvard University.

Generation of *Cbx2*^{23KRA} targeting construct

The targeting construct was assembled from the following components into a pUC19 vector backbone: gBlock encoding the K-to-A and R-to-A mutations (Integrated DNA Technologies, IDT); homologous arms PCR amplified from a BAC clone (RP23-272P15); thymidine kinase selection cassette PCR amplified from the plasmid PL253 (ATCC, PTA-4998); loxP-flanked neomycin selection cassette amplified from a standard plasmid. We verified the construct by sequencing.

Production of sgRNA and donor DNA for CRISPR-mediated gene targeting

sgRNAs (sgRNA-L: GCCTTCTGTTCTGGGGGTAGAGG; sgRNA-R: GAGACCCGTCAGCCTGGC CAAGG) were designed using Zinc Finger Targeter (ZiFiT)[37,38], synthesized as DNA oligonucleotides (IDT), annealed, phosphorylated and cloned into pX335(Addgene plasmid # 42335). T7 promoter was added to sgRNA template by PCR amplification using primers CLO94/95 (Appendix, Table S3). The T7-sgRNA PCR product was gel purified and used as the template for *in vitro* transcription using MEGAshortscript T7 kit (Life Technologies). sgRNAs were purified using MEGAclear kit (Life Technologies) and eluted in RNase-free water.

The donor DNA was assembled from the following components into a pUC19 vector backbone: gBlock encoding the K-to-A and R-to-A mutations (IDT) and homologous arms

PCR amplified from a BAC clone (RP23-272P15) using primer pairs CLO84/85 and CLO86/87. The assembled donor was purified using UltraClean Endotoxin Removal (MO Bio) and eluted in nuclease-free water.

The Cas9n mRNA was purchased from SystemBiosciences (CAS504A-1).

Genotyping by PCR

Genomic DNA was isolated from mouse toe or tail tip tissues using NucleoSpin Tissue (Machery-Nagel), and genotyped by PCR amplification using primers indicated in the figures, followed by gel electrophoresis analysis of the products. Primer sequences are listed in Appendix Table S3. For RFLP analysis, the PCR products were purified, treated with HindIII and XmaI for at least 3 hours at 37degG, heat inactivated, and analyzed by gel electrophoresis.

Genotyping by melt curve analysis

Exon 5 of *Cbx2* is amplified by real-time PCR using the primer pairs gDet23F/gDet24 and iTaq Universal SYBR Green Supermix (BioRad #1725120). The melt curve profiles were generated by increasing the temperature from 85.0 to 91.0 degC in 0.1 degC increments, and analyzed using Bio-Rad's CFX Manager™ software.The PCR products from wild type and *Cbx2* mutant alleles give different profiles due to differences in the nucleotide sequence. We verified the results from this method by cloning and sequencing the PCR products.

Western blot

Lysates were prepared by lysing whole embryos in RIPA buffer (50mM Tris-HCl pH 8.0, 0.1% SDS, 0.5% sodium deoxycholate, 1% NP40, 150mM NaCl), separated on 4-12%

bis-Tris gels, transferred onto nitrocellulose membranes and blotted for CBX2 and CBX8 using the antibodies listed in Appendix Table S1. H3 levels were used as loading controls. Relative protein levels were quantified using ImageJ.

Protein Purification

Flag-tagged CBX2 variants and untagged Ring1b were expressed and purified as previously described[9]. Briefly, the coding sequences were inserted into pFastBac1. Baculovirus was generated following the Bac-to-Bac® system procedures (Thermofisher). The proteins were by co-infecting Sf9 cells. Sf9 nuclear extracts were generated using the Dignam protocol, flash-frozen, and stored at -80°C until needed[39]. To purify proteins, nuclear extracts were thawed in cold water and then incubated with M2-agarose (Sigma Aldrich A2220) at 4°C with gentle rocking for 4 hrs. The M2-agarose was collected by centrifuging at 250 g for 5 min, washed in batch first with BC buffer (20mM HEPES 7.5, 0.2mM EDTA, 20% glycerol, 1mM DTT,) containing 500mM KCl, then 300mM KCl. Proteins were eluted in BC buffer plus 300mM KCl containing 0.4 mg/ml flag peptide and quantified using standard techniques.

Nucleosome Remodeling and Inhibition Assays

A plasmid containing 10 5S nucleosome positioning sequences and a unique centrally located HhaI restriction enzyme site (pG5E4) has been previously described[40]. The 2.5 kb Asp718/ClaI G5E4 fragment was digested, purified, and end labeled with Cy5. Chromatin arrays were assembled using HeLa core octamers and a salt-gradient assembly protocol, and were typically 80-85% assembled by observing background cutting with

HhaI[41]. Reactions were done essentially as described in . PcG proteins were incubated with 0.5-3nM nucleosomes for 30 min at 30°C, and then 100ng of purified human SWI/SNF and 8 units of HhaI were added to a volume of 20µl. The reactions were allowed to proceed for 1 hr at 30°C, and then 5µl of stop buffer was added and reactions incubated an additional 30 min at 55°C. Final reaction buffer conditions were as follows: 16 mM HEPES 7.5, 80mM KCl, 2mM $MgCl_2$, 1mM $MnCl_2$, 2mM ATP, 2.5 µg BSA, 8% glycerol, and 2 mM DTT. Stop buffer consists of 10mM Tris 7.5, 70mM EDTA, 1% SDS, 0.1% Orange G, and 1.5 mg/ml Proteinase K. After completion, reactions were separated on 1% agarose TAE gels and scanned and quantified using a Typhoon phosphoimager and ImageQuant TL software (GE Healthcare). Graphs were generated using KaleidaGraph graphing software (Synergy).

Skeletal analysis

Skeletal whole mounts of P15-17 mice were performed on completely eviscerated animals with the skin removed. Carcasses were incubated in 96% ethanol for 5 days, and transferred to acetone for 2 days. Staining was performed in 0.005% alizarin red S, 0.015% alcian blue 8GS in 5% acetic acid, 5% H20, and 90% ethanol for 3 days at 37degC or 5 days at room temperature. Samples were rinsed with water and cleared by sequential incubation in 1% KOH (2 days), 0.8% KOH and 20% glycerol, 0.5% KOH and 50% glycerol, and 0.2% KOH and 80% glycerol. Cleared skeletons were stored in 100% glycerol

Histological analysis

The animals were fixed in Bouin's solution, and various organs were sectioned and stained with hematoxylin and eosin for examination. The Rodent Histopathology Core at Dana-Farber/Harvard Cancer Center performed the sectioning and slide analyses.

REFERENCES

1. Lewis, E B. New mutants: Reports of P. Lewis, *Drosophila Information Service* **21,** 69 (47AD).

2. Jürgens, G. A group of genes controlling the spatial expression of the bithorax complex in Drosophila. *Nature* **316,** 153–155 (1985).

3. B, L. E. A gene complex controlling segmentation in Drosophila. *Nature* **276,** 565–570 (1978).

4. G Struhl, M. A. Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila. *EMBO J* **4,** 3259 (1985).

5. Moazed, D. & O'Farrell, P. H. Maintenance of the engrailed expression pattern by Polycomb group genes in Drosophila. *Development* **116,** 805–810 (1992).

6. Aranda, S., Mas, G. & Di Croce, L. Regulation of gene transcription by Polycomb proteins. *Sci Adv* **1,** e1500737 (2015).

7. van der Lugt, N. M. *et al.* Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes & Development* **8,** 757–769 (1994).

8. Akasaka, T. *et al.* A role for mel-18, a Polycomb group-related vertebrate gene, during theanteroposterior specification of the axial skeleton. *Development* **122,** 1513–1522 (1996).

9. Grau, D. J. *et al.* Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes & Development* **25,** 2210–2221 (2011).

10. Bernstein, E. *et al.* Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin. *Mol. Cell. Biol.* **26,** 2560–2569 (2006).

11. Bezsonova, I. *et al.* Ring1B Contains a Ubiquitin-Like Docking Module for Interaction with Cbx Proteins . *Biochemistry* **48,** 10542–10548 (2009).

12. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nat Cell Biol* **441,** 349–353 (2006).

13. Morey, L. *et al.* Nonoverlapping Functions of the Polycomb Group Cbx Family of Proteins in Embryonic Stem Cells. *Cell Stem Cell* **10,** 47–62 (2012).

14. Vincenz, C. & Kerppola, T. K. Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. *PNAS* (2008).

15. Creppe, C., Palau, A., Malinverni, R., Valero, V. & Buschbeck, M. A Cbx8-Containing Polycomb Complex Facilitates the Transition to Gene Activation during ES Cell Differentiation. *PLoS Genet.* **10,** e1004851 (2014).

16. Senthilkumar, R. & Mishra, R. K. Novel motifs distinguish multiple homologues of Polycomb in vertebrates: expansion and diversification of the epigenetic toolkit. *BMC Genomics* **10,** 549 (2009).

17. Ren, X., Vincenz, C. & Kerppola, T. K. Changes in the Distributions and Dynamics of Polycomb Repressive Complexes during Embryonic Stem Cell Differentiation. *Mol. Cell. Biol.* **28,** 2884–2895 (2008).

18. Tatavosian, R. *et al.* Distinct Cellular Assembly Stoichiometry of Polycomb Complexes on Chromatin Revealed by Single-Molecule Chromatin Immunoprecipitation Imaging. *Journal of Biological Chemistry* (2015). doi:10.1074/jbc.M115.671115

19. Kato, Y., Koseki, H., Vidal, M., Nakauchi, H. & Iwama, A. Unique Composition of Polycomb Repressive Complex 1 in Hematopoietic Stem Cells. *International Journal of Hematology* **85,** 179–181 (2007).

20. Klauke, K. *et al.* Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. *Nat Cell Biol* **15,** 353–362 (2013).

21. van den Boom, V. *et al.* Nonredundant and locus-specific gene repression functions of PRC1 paralog family members in human hematopoietic stem/progenitor cells. *Blood* **121,** 2452–2461 (2013).

22. Luis, N. M. *et al.* Regulation of Human Epidermal Stem Cell Proliferation and Senescence Requires Polycomb- Dependent and -Independent Functions of Cbx4. *Cell Stem Cell* **9,** 233–246 (2011).

23. Ma, R.-G., Zhang, Y., Sun, T.-T. & Cheng, B. Epigenetic regulation by polycomb group complexes: focus on roles of CBX proteins. *J Zhejiang Univ Sci B* **15,** 412–428 (2014).

24. Coré, N. *et al.* Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice. *Development* **124,** 721–729 (1997).

25. Katoh-Fukui, Y. *et al.* Male-to-female sex reversal in M33 mutant mice. *Nature* **393,** 688–692 (1998).

26. Forzati, F. *et al.* CBX7 is a tumor suppressor in mice and humans. *J. Clin. Invest.* (2012). doi:10.1172/JCI58620DS1

27. Cbx4<tm1.2Gxu> Targeted Allele Detail MGI Mouse (MGI:5489932). http://www.informatics.jax.org/allele/MGI:5489932.

28. Cbx8<tm1.1Hko> Targeted Allele Detail MGI Mouse (MGI:5306630). http://www.informatics.jax.org/allele/MGI:5306630

29. Cbx6<tm1a(EUCOMM)Wtsi> Targeted Allele Detail MGI Mouse (MGI:4441849). http://www.informatics.jax.org/allele/MGI:4441849.

30. Biason-Lauber, A., Konrad, D., Meyer, M., deBeaufort, C. & Schoenle, E. J. Ovaries and Female Phenotype in a Girl with 46,XY Karyotype and Mutations in the CBX2 Gene. *The American Journal of Human Genetics* **84,** 658–663 (2009).

31. Liu, B. *et al.* Cbx4 regulates the proliferation of thymic epithelial cells and thymus function. *Development* **140,** 780–788 (2013).

32. Tan, J. *et al.* CBX8, a polycomb group protein, is essential for MLL-AF9-induced leukemogenesis. *Cancer Cell* **20,** 563–575 (2011).

33. Katoh-Fukui, Y. *et al.* Male-to-female sex reversal in M33 mutant mice. *Nature* **393,** 688–692 (1998).

34. Isono, K. *et al.* SAM Domain Polymerization Links Subnuclear Clustering of PRC1 to Gene Silencing. *Developmental Cell* **26,** 565–577 (2013).

35. Kessel, M. & Gruss, P. Homeotic transformations of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid. *Cell* **67,** 89–104 (1991).

36. Richardson, L. *et al.* EMAGE mouse embryo spatial gene expression database: 2014 update. *Nucleic Acids Research* **42,** D835–44 (2014).

37. Sander, J. D., Zaback, P., Joung, J. K., Voytas, D. F. & Dobbs, D. Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Research* **35,** W599–605 (2007).

38. Sander, J. D. *et al.* ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Research* **38,** W462–8 (2010).

39. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Research* **11,** 1475–1489 (1983).

40. Steger, D. J., Eberharter, A., John, S., Grant, P. A. & Workman, J. L. Purified histone acetyltransferase complexes stimulate HIV-1 transcription from preassembled nucleosomal arrays. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 12924–12929 (1998).

41. Côté, J., Utley, R. T. & Workman, J. L. in *Methods in Molecular Genetics* **6,** 108–128 (Elsevier, 1995).

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

We tested the hypothesis that the *in vitro* compaction activity of CBX2 is required for PRC1 function in *vivo.* In Chapter 2, we showed that mutations that disrupt CBX2 compaction activity *in vitro* prevented CBX2 from repressing its target genes in mouse embryonic stem cells. In Chapter 3, we showed that the same mutations and a subset of them caused posterior transformations in mice, which is the classic PcG phenotype. Together, our work indicates that the *in vitro* compaction activity of PRC1, as conferred by CBX2, is essential for its function *in vivo*.

In the following sections, we discuss what may be interesting future research directions following the outcome of our study.

Characterizing compacted chromatin *in vivo*

It is conceptually simple to see how PRC1's chromatin compaction activity can reduce DNA accessibility to transcriptional activators and the transcription machinery, and thereby repress gene expression. However, as discussed in Chapter 1, there is no strong evidence that indicates DNA accessibility is markedly reduced by PRC1 function. Moreover, unlike heterochromatin, PcG bound regions are more dynamic than previously perceived[1-3]. This leads one to ask what are the characteristics of PcG compacted chromatin *in vivo.*

In the previous studies, the assays for accessibility were either performed on single transgenes, or on one or two endogenous PRC1 targets [4-7]. The results appeared to be

sensitive to the nature and genomic location of the locus under study. This suggests that the examination of single targets is insufficient for providing information about a general characteristic of PRC1 bound regions. More recently, a genome-wide study, which used the methylase M.SssI on fly S2 cells, showed that PcG-associated regions are collectively less accessible than non- PcG-associated regions, even though there is considerable variation among individual regions.[8]

In view of our results presented here, which indicate that chromatin compaction at the nucleosomal level is an important activity of PRC1 function, it seems warranted that we re-visit the concept of chromatin accessibility using different technologies or approaches. This will inform us on a fundamental nature of chromatin that is modulated for transcriptional regulation. Several methodologies have been developed but have not been tested for their utilities to detect PcG-dependent changes in the genome. We propose to apply them to our *Cbx2* mutants which are deficient in the compaction activity. Examples of these include micrococcal nuclease (MNase) titration and ATAC-seq. ATAC-seq has been described extensively elsewhere (for examples, see Ref.9), and hence will not be described here. We only briefly describe MNase titration below.

MNase enzyme preferentially cuts linker DNA on chromatin, thereby producing DNA fragments which correspond to nucleosomal DNA that run in a signature 150bp laddering pattern on gel electrophoresis analysis. It was observed that different amounts of MNase used for chromatin digestion releases distinct populations of nucleosomal DNA, which presumably reflect different degrees of nucleosome occupancy in the genome. Specifically, at low MNase concentrations, DNA around labile nucleosomes is recovered while at higher MNase concentrations, DNA around refractory nucleosomes is recovered. The behavior of a DNA region in response to digestion with a series of MNase concentration (hence, titration) can be measured in a metric called MNase accessibility score (MACC) (Tolstorukov M. *et.al,*

manuscript under review). Work in our lab has shown that MACC is low (i.e. low accessibility) for nucleosomes in PcG-bound regions in fly S2 cells and human K562 cells, compared to genome average. Importantly, the low MACC score at PcG-bound regions in flies is alleviated by PcG knockout mutations, suggesting that MACC could be used as readout for PRC1-dependent compaction. It would be interesting to determine whether mutation in the CBX2 compaction domain would alter MACC in the mouse genome.

Elucidating the CBX2-nucleosome interaction

How the compaction domain of CBX2 interfaces with chromatin to achieve compaction is unknown. The fact that it is intrinsically disordered and has high positive charge presents two possibilities.

The first model involves non-directed electrostatic interactions between the positively charged domain and the negatively charged DNA. The configuration a chromatin fiber is highly sensitive to the ionic condition of its surroundings; it spontaneously folds into higher order structures in the presence of high cation concentrations[10]. It is conceivable that the compaction domain of CBX2 behaves like a series of regulated cations that is brought to the vicinity of chromatin by PRC1 to induce chromatin folding.

The second model involved specific interactions between functional groups on the compaction domain and chromatin. It is known that a region with intrinsic disorder may become folded upon interaction with other proteins. This could be the case for the compaction domain in CBX2, where upon its contact with chromatin becomes ordered and presents certain amino acid sides chains for interaction with the nucleosomes. In fact, structural studies with other nucleosome binding proteins suggest that this could be a common feature of protein-nucleosome interactions. The protein-nucleosome interfaces frequently involve arginine side chains from the protein making contacts with the acidic

patch on the nucleosome face[11-15]. This led to the 'arginine anchor' hypothesis for protein-nucleosome interactions.

For our CBX2 compaction mutants, we mutated both lysine (K) and arginine (R) residues to alanine, which removed both positive charge and the side chains that provided the charge; hence, we are unable to distinguish between the two models of interaction with our mutants.

To separate charge from functional group, one can generate CBX2 variants with all R-to-K and all K-to-R amino acid substitutions in the compaction domain and compare their *in vitro* activity levels. These mutations do not alter the overall charge of the protein but modify the functional groups available. Through systematically comparing a series of CBX2 variants with R-to-K and K-to-R changes at different residues, one may be able to identify the critical arginine residue(s), if indeed the arginine anchor hypothesis applies to CBX2. Alternatively, one may attempt to obtain information about CBX2-nucleosome interaction via structural approaches such as X-ray crystallography or high-resolution cryo-electron microscopy.

Propagation of the repressed state

PcG-mediated repression needs to be propagated through cellular division in order to be inherited in daughter cells. This entails transmitting the associated chromatin states through mitosis, or rapidly restoring them in the following G1 phase.

The kinetics of CBX2 binding on interphase chromatin and mitotic chromosomes are very different. In interphase, CBX2's association with chromatin is dynamic, where its chromatin bound and nucleoplasmic fractions interchange rapidly[16]. In mitosis, however, CBX2 becomes immobilized on the chromosomes, along with Ring1B, Mel-18 and Phc1, which are subunits of the canonical PRC1[17]. Similar observations have been reported for the

fly system, where sub-fractions of PC, PH and PSC stay immobilize on chromosomes during mitosis[18,19]. The stable binding of PRC1 components on mitotic chromosomes suggest that they may be retained to 'bookmark' key sites so that PcG repressed regions can be rapidly restored following cell division.

Among the five mammalian PC homologs CBX2/4/6/7 and 8, only CBX2 immobilizes on mitotic chromosomes[17]. Its recruitment to chromosomes is independent of the other PRC1 subunits and PRC2. Conversely, it is required for their recruitment. The chromodomain of CBX2 is necessary for its own recruitment, while its C-box domain is necessary for its immobilization. The chromodomain and C-box domains are conserved in the other CBX proteins, but are insufficient for their retentions on chromosomes. This indicates that additional regulatory factors, and/or the non-homologous regions of CBX2 must also contribute to its unique binding properties. The compaction domain which we showed to be important for PcG function is a unique feature of CBX2; it will be interesting to determine whether it is required for CBX2 association with mitotic chromosomes.

The expression levels of *Cbx2* in whole embryos and adult tissues appear to correlate with the rate of cellular division. For example, CBX2 level is generally lower in adult tissues compared to embryos, consistent with the adult tissues having lower mitotic indexes[20]. Among the adult tissues, CBX2 is high in organs with active cell proliferation such as the bone marrow, liver and intestines, but is negligible in post-mitotic tissues such as the brain, heart and muscle[21]. From these observations, together with CBX2 immobilization on mitotic chromosomes, one may hypothesize that CBX2, in the context of PRC1, has a fundamental role in propagating cellular identities in rapidly dividing cells. Therefore, increased molecular and mechanistic understanding of CBX2 may yield promising insights into the mechanisms of epigenetic inheritance.

REFERENCES

1. Mito, Y., Henikoff, J. G. & Henikoff, S. Genome-scale profiling of histone H3.3 replacement patterns. *Nat Genet* **37,** 1090–1097 (2005).

2. Mito, Y., Henikoff, J. G. & Henikoff, S. Histone Replacement Marks the Boundaries of cis-Regulatory Domains. *Science* **315,** 1408–1411 (2007).

3. Deal, R. B., Henikoff, J. G. & Henikoff, S. Genome-Wide Kinetics of Nucleosome Turnover Determined by Metabolic Labeling of Histones. *Science* **328,** 1161–1164 (2010).

4. Schloßherr, J., Eggert, H., Paro, R., Cremer, S. & Jack, R. S. Gene inactivation in Drosophiia mediated by the Polycomb gene product or by position-effect variegation does not involve major changes in the accessibility of the chromatin fibre. *Molec. Gen. Genet.* **243,** 453–462

5. Boivin, A. & Dura, J.-M. In vivo chromatin accessibility correlates with gene silencing in Drosophila. *Genetics* (1998).

6. McCall, K. & Bender, W. Probes of chromatin accessibility in the Drosophila bithorax complex respond differently to Polycomb-mediated repression. *EMBO J* **15,** 569 (1996).

7. Fitzgerald, D. P. & Bender, W. Polycomb group repression reduces DNA accessibility. *Mol. Cell. Biol.* **21,** 6585–6597 (2001).

8. Bell, O. *et al.* Accessibility of the Drosophila genome discriminates PcG repression, H4K16 acetylation and replication timing. *Nat Struct Mol Biol* **17,** 894–900 (2010).

9. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* **10,** 1213–1218 (2013).

10. Korolev, N. *et al.* Electrostatic Origin of Salt-Induced Nucleosome Array Compaction. *Biophysical Journal* **99,** 1896–1905 (2010).

11. McGinty, R. K., Henrici, R. C. & Tan, S. Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature* **514,** 591–596 (2014).

12. Barbera, A. J. *et al.* The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science* **311,** 856–861 (2006).

13. Makde, R. D., England, J. R., Yennawar, H. P. & Tan, S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467,** 562–566 (2010).

14. Kato, H. *et al.* A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. *Science* **340,** 1110–1113 (2013).

15. Armache, K.-J., Garlick, J. D., Canzio, D., Narlikar, G. J. & Kingston, R. E. Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science* **334,** 977–982 (2011).

16. Ren, X., Vincenz, C. & Kerppola, T. K. Changes in the Distributions and Dynamics of Polycomb Repressive Complexes during Embryonic Stem Cell Differentiation. *Mol. Cell. Biol.* **28,** 2884–2895 (2008).

17. Zhen, C. Y., Duc, H. N., Kokotovic, M., Phiel, C. J. & Ren, X. Cbx2 stably associates with mitotic chromosomes via a PRC2- or PRC1-independent mechanism and is needed for recruiting PRC1 complex to mitotic chromosomes. *Molecular Biology of the Cell* **25,** 3726–3739 (2014).

18. Fonseca, J. P. *et al.* In vivo Polycomb kinetics and mitotic chromatin binding distinguish stem cells from differentiated cells. *Genes & Development* **26,** 857–871 (2012).

19. Follmer, N. E., Wani, A. H. & Francis, N. J. A Polycomb Group Protein Is Retained at Specific Sites on Chromatin in Mitosis. *PLoS Genet.* **8,** e1003135 (2012).

20. Pearce, J. J., Singh, P. B. & Gaunt, S. J. The mouse has a Polycomb-like chromobox gene. *Development* **114,** 921–929 (1992).

21. Hashimoto, N. *et al.* RAE28, BMI1, and M33 Are Members of Heterogeneous Multimeric Mammalian Polycomb Group Complexes. *Biochem. Biophys. Res. Commun.* **245,** 356–365 (1998).

APPENDIX

Table S1 Antibody list
The amounts used for the various applications are indicated.

Antibodies	Company	Catalog Number	ChIP	IP	Western-blot
CBX2	Bethyl Laboratories Inc	A302-524A	4ug		
CBX2 (C-18)	Santa-Cruz	sc-19297			1/500
CBX7	Abcam Inc	ab21873	4ug		
CBX7	EMD Millipore	07-981			1/2000
CBX7 (G-3)	Santa-Cruz	sc-376274			1/500
CBX8	Bethyl Laboratories Inc	A300-882A			1/1000
H2AK119Ub	Cell Signaling Technology	8240S	4ug		
H3	Abcam Inc	ab1791			1/10,000
H3K27me3	EMD Millipore	07-449	4ug		
HA	Covance Research Inc	MMS-101P			1/2000
HA	Abcam Inc	ab9110			1/2000
Phc1	Active Motif	39723			1/2000
Rabbit IgG	Abcam Inc	ab37415		4ug	
Ring1B	Bethyl Laboratories Inc	A302-869A	4ug	4ug	
Ring1B	Active Motif	39663			1/2000
TBP	Abcam Inc	ab818			1/2000

Table S2 Taqman probe list

Gene	Assay ID
Sall3	Mm01265835_m1
Wt1	Mm01337048_m1
Six1	Mm00808212_m1
Ebf1	Mm01288946_m1
Mllt10	Mm00487708_m1
B2m	Mm00437762_m1
Hprt	Mm00446968_m1
Pou5f1	Mm00658129_gH
Nanog	Mm02384862_g1
Actb	Mm00607939_s1

Table S3 Primer list

ChIP primers	
Mllt10_F	CAATGCCTCGTCACAGCTAATA
Mllt10_R	GCCTCCTGTGAAAGTGAAAGA
Sall3_F	CAGAGATGGAACTGGTGCTTTA
Sall3_R	GCCCAGAGTACAGTGTGTTAG
Wt1_F	GGAAGTGTTGGGAATCTCTCTT
Wt1_R	CTCCAGGGATTTGGTTCGTATC
Six1_F	CACTGTCCCGAGAAACATAGAG
Six1_R	GTATGAGGAGAGGATAGGGATAGG
Ebf1_F	CTCCTGACTTCTGTGTGGTTT
Ebf1_R	CTGAGTCCTGGTTACACATAGC
Gata6-F	GCTCCTTTCCCAGAGCGTTGAAT
Gata6-R	CCCTCCTTCCAAATTAAGCCC
Hoxa11-F	GGAAGTAAGCAGAAAGATACAGGGAAGG
Hoxa11-R	TTTGTCAATAATCCGCGCTGTCCG
Sox17-F	TTACTTGTGGCATTTGTGGCTGGC
Sox17-R	CAGCAGTGTGAGTGGGCCATATTT
Gene desert_F	CCACTCTTCTTATAGGACCCTTTG
Gene desert_R	CCTGTCTACCTGTTCTTTACATTCT
Gata1_F	GATCACCCTGAACTCGTCATAC
Gata1_R	TTTGGGAATCAAGACTGACCTG

Cloning and genotyping primers	
CLO75	GGAGTCTAGGTCGCTGTCGT
CLO84	GTGAATTCGAGCTCGGTACCCGCTCTGCGGGGCTAACCGC
CLO85	AGAGGACGAACTGCTGGATTTGGA
CLO86	CATGAGCCAGAGTCAGGTTCAGGC
CLO87	TTGCATGCCTGCAGGTCGACTCAGCACTAGACTTATCTCC
CLO94	TTAATACGACTCACTATAGG
CLO95	AAAAGCACCGACTCGGTGCC
gDet23	GAACCAGATGCGCCATCCAAATC
gDet24	TCATGTAGTGTACGATGGAGC
gDet28	GGTGGTCAGACTAGGAACACAGAAAT
gDet33	CCGATTCGCAGCGCATCGCCTTCTATCGCC
gDet39	GAATCAACGCCATAACTTCG
seqP28	CCCCTATTGGAGCTAGAGCA
seqP37	AGTTTGCCTGTGGGTTTGAG
seqP55	AAAGGCACCAAATTGGGATT